DOCUMENT RESUME

ED 299 706 EA 020 447

TITLE	Creating Responsible and Responsive Accountability Systems. Report of the OERI State Accountability Study Group.
INSTITUTION	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	PIP-88-808
PUB DATE	Sep 88
NOTE	78p.
PUB TYPE	Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Academic Achievement; *Accountability; Attendance; *Data Collection; Elementary Secondary Education; *Performance; *Public Education; *State School District Relationship; *State Standards
IDENTIFIERS	*Indicators

ABSTRACT

Most states pursue accountability in public education by collecting and reporting school data, but types of data collected, and how those data are reported and used in holding schools accountable, differ from state to state. This report is designed for state policymakers desiring to refine their existing accountability systems. The report's six sections raise basic design issues and provide research- and experience-based guidelines without providing definitive answers. Following an executive summary and introduction, section 2 describes the basic elements of a performance accountability system. Section 3 describes state accountability system variations related to primary responsibility; linkage with other state policies; indicator types; data collection, organization, and reportage levels; reporting mechanisms; and school comparison methods. Section 4 discusses uses and consequences of performance accountability systems. Local educators' actual use of these systems does not necessarily match state policymakers' assumptions. State-controlled and locally controlled systems are briefly compared. Section 5 treats unresolved dilemmas, including balancing oversight and improvement, determining the appropriate accountability level, balancing statewide comparability with local ownership, expanding the alternatives to traditional standardized tests, making fair comparisons, and ensuring adequate capacity. In the final section, recommendations are given for designing better systems and for using accountability data to improve low-achieving schools and encourage high-performing or significantly improving schools. Included are 14 references the three appendices listing state accountability study group members and additional resource persons and providing tables of educational indicators. (MLH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SEP 8 3 1994

ED 369206

Creating Responsible and Responsive Accountability Systems

U.S. DEPARTMENT OF EDUCATION

Programs for the Improvement of Practice
Office of Educational Research and Improvement
U.S. Department of Education

LB 2822.82 .H47

Creating Responsible and Responsive Accountability Systems

Report of the OERI State Accountability Study Group

Programs for the Improvement of Practice
Office of Educational Research and Improvement
U.S. Department of Education

U.S. Department of Education
Lauro F. Cavazos
Secretary

Office of Educational Research and Improvement
Chester E. Finn, Jr.
Assistant Secretary

Programs for the Improvement of Practice
Milton Goldberg
Director

Information Services
Ray Fields
Director

September 1988

Preface

More and more citizens want to know in concrete terms what they are receiving in return for their hard-earned tax dollars invested in public education. Governors, legislators, State leaders, and blue-ribbon committees are increasingly being called upon to provide more financial support for public schools. They naturally want to be assured that this support yields an adequate return.

As a result of these demands, concern about student performance, school results, and accountability for outcomes and improvement continues to grow. This report attempts to provide alternatives to citizens, businesspeople, parents, policymakers, and educators who expect increased school accountability and productivity.

The public may be displeased at times by the less than enthusiastic response educators might give to initial cries for accountability in education. Educator displeasure is often not about the principle that schools should be held accountable. Rather it stems from their legitimate concern about the accountability measures selected. These measures may focus education too narrowly, may cause schools to pay attention to rote learning only, may encourage teaching to test items, or may be inappropriately used. This report discusses such matters and offers options for providing school accountability in a way that could minimize a number of frequently expressed concerns.

Unfortunately, or fortunately, from some points of view, there is no one, tried and tested solution to school accountability. There are, however, several principles upon which some agreement can be reached. Our Study Group, staff, and researchers reached agreement on these principles after careful study of accountability measures in the 10 States represented in the group; fieldwork at schools in 4 States; analysis of a 50-State survey; review of the experience of 5 State-based, national organizations; and research in this area. As a result, the background information, findings, and recommendations in this report represent informed judgment from an unusually diverse mixture of sources and people working together.

Good public education for all youth is critical to meet the societal and economic challenges facing America. A thoughtful approach to school accountability is one ingredient in providing that good education.

The reader is encouraged to use the suggested accountability measures not with "an ax to grind" but as "a field to sow."

Terry Peterson
Chair, Office of Educational Research and Improvement
State Accountability Study Group

Foreword

As in business and government, in education, reliable information is one key to better performance. This belief has guided many recent efforts of the U.S. Department of Education. From *American Education: Making It Work* and *What Works: Research About Teaching and Learning* to a host of other projects, products, and activities, better information about the condition of schools has been made available to the public.

While the Department has been working at the national level, State education leaders across the country have been working to improve the caliber of schools and build enduring systems for collecting and reporting sound information on their performance. We knew that States like California, Illinois, and South Carolina were developing public "report cards" on the performance of individual schools throughout their States. We knew that New Jersey, South Carolina, and Georgia were pioneering a concept of educational "bankruptcy" for schools and school systems found to perform chronically at substandard levels. We knew that other States were devising rewards and assistance to improve school performance and that policy was changing faster than practically anyone could easily keep track of.

Accordingly, in the summer of 1987, the Department convened a study group of individuals from 10 States and five national organizations, all of them knowledgeable about promising new trends in State educational accountability. Their charge was to help us understand these developments and, based on their collective experience, to offer recommendations to other States. In less than one year's time they were able to do so. This book is the report of their work.

This OERI State Accountability Study Group, as we called it, was made up of experienced, State-based people from governors' offices, State legislatures (or their staffs), departments of education, and local school districts. They were joined by staff with similar but national perspectives—from the National Governors' Association (NGA), the National Conference of State Legislatures (NCSL), the Council of Chief State School Officers (CCSSO), the National Association of State Boards of Education (NASBE), and our own Department staff. Two policy analysts, Lorraine McDonnell and Jeannie Oakes of the RAND Corporation, as part of their work with the Center for Research on Educational Policy (CPRE), were the Study Group members who drafted the report. The group was chaired by Terry Peterson of South Carolina, currently executive director of the State's business- education oversight committee and previously Governor Dick Riley's chief education aide.

Meeting in July and November 1987 and again in February 1988, the Study Group reported on the practices of their State accountability systems, how they developed over time, and the problems and successes the policies encountered. After the first meeting, the CCSSO was commissioned to collect information on current practices of all 50 States and the District of Columbia concerning key elements of State educational accountability. This is the source of much of the 50-State data found in this report. The exact numbers—how many States do what—have been changing rapidly, even

during the brief life of the Study Group. Nonetheless, this report presents the best analysis and most current information available on these evolving policies and practices.

If we are successful, this report will encourage its own obsolescence, as States move to refine the data they collect, the audiences to whom they report it, the format and provision for its interpretation, and most of all, the skill and wisdom with which the States use that information to improve their schools. As such changes occur, our purpose is to assist those who are building accountability systems by offering them an overview of the nature of the technical and political issues involved and the best advice from experienced colleagues.

The creation of sound State accountability systems is one of the most potent of the recent education reforms of the States. The systems will produce comprehensive information about schools for years to come. At their best, they will allow policymakers, educators, and the public to know how well their schools are now doing and how to help them do better in the future. As this report makes clear, this is no easy task. In many communities, it involves new definitions of who is going to be held accountable to whom for what. These are potentially fractious issues whose resolution lies in clarity of purpose, mutuality of trust, and evolution of consensus on the point that solid information and well wrought accountability mechanisms are essential for the success of schools.

Chester E. Finn, Jr.
Assistant Secretary
and Counselor to the Secretary

Executive Summary

The majority of States pursue accountability in public education by collecting and reporting data about schools—student attendance and test performance, resources spent, the types of programs offered, and the like. But the kinds of data collected, and how those data are reported and used in holding schools accountable, differ from State to State. This report, generated by the OERI State Accountability Study Group, is designed for State policymakers who are interested in refining their existing educational accountability systems. The report raises basic design issues and provides guidelines based on research and the experience of other States. However, it does not supply definitive answers. Each State must design an accountability strategy that supports its own educational goals and needs and is consistent with its own political traditions.

Basic Elements of a Performance Accountability System

A performance accountability system is a set of indicators or statistics that provides information about how well schools are performing. Data from the system should allow policymakers to compare performance over time, against standards, and with comparable educational entities (for example, States with other States, schools with other schools). By their choice of indicators, policymakers determine *who* will be held accountable, *for what*, and *to whom*.

To produce information that is reliable, fair, and useful for improving schools, the indicators in the accountability system should

- Measure the central features of schooling;
- Measure what is actually being taught or considered important for students to know;
- Provide information that is policy relevant;
- Focus on the school site;
- Allow for fair comparison; and
- Maximize the usefulness of the data collected and minimize the burden of collecting it.

An accountability system is more effective and more likely to be used if it relies on a model that represents the entire schooling enterprise.

State Performance Accountability Systems: Description and Trends

All 50 States and the District of Columbia collect performance data about schooling. Although the systems vary considerably, they can be classified along two major dimensions: 1) who is primarily responsible for the system and 2) whether the system is linked with other State policies. The dominant pattern, found in 21 States, is one in which the State has primary responsibility for deciding what data to collect and how to organize and report them and what State-level policy action, based on the data, will improve education. No State system can accurately be considered a "pure" type, however, and States with the same general type can have dissimilar systems.

States must also specify:

What the indicators will be:

- Student performance data?
- School resources and processes?
- Background data on students and communities?

The level at which data will be collected:

- State?
- District?
- School?
- Classroom?
- Student?

The level at which data will be organized and reported:

- State?
- District?
- School?

The mechanisms for reporting:

- Reports to the public and media?
- Reports to parents?
- Data made available on request?

How schools will be compared:

- With all other schools?
- With similar schools?
- With their own past performance?

Each of these decisions shapes the content and form of a State's accountability system. Whether the resulting system is effective, feasible, and acceptable to practitioners and policymakers depends on the State's political context and existing policy precedents. Consequently, no best design can be prescribed for all States, and, not surprisingly, no two States have made identical decisions in all areas.

Uses and Consequences of Performance Accountability Systems

In designing their systems, State policymakers make certain assumptions about how accountability data will be used—both for their own purposes and by local districts and schools. However, the way the systems are actually used depends on how local educators choose to incorporate them and the information they generate into school practice.

The investigation of the ways in which accountability systems are influencing schools show mixed results. Although these systems are powerful policy tools that have significantly changed school-level planning and teaching activities, many respondents feel that they motivate teachers and administrators to focus narrowly on tests.

State-controlled systems differ from locally controlled systems. With State-controlled accountability, local districts and schools appear to take the systems seriously, and have changed their behavior in order to improve performance. However, that performance has been narrowly defined by the focus and scope of State tests. Many school personnel therefore do not consider the systems useful or valid measures of their schools' performance. Instead, they respond to the systems because they feel pressure for their students to do well. Although local superintendents and principals recognize the need to use other indicators as well in judging a district's or school's performance, they see the State as being primarily concerned with test scores; consequently, they have channeled much of their energy to improving performance on that one indicator.

Locally controlled systems have both more and less effect than State-controlled systems. In districts where school officials are committed to using accountability data to improve schools, the locally controlled system influences school practice more than a State-operated one. However, whether this process occurs or not depends entirely on local priorities and capacity. Some districts use the data extensively to inform practice, while others do little more than meet State testing requirements.

Unresolved Dilemmas

Policymakers face a number of major dilemmas in developing or modifying a State accountability system, including:

Balancing oversight and improvement.

Policymakers can ensure that an accountability system provides useful information for improving schools, not just monitoring them, by carefully choosing the types of data that are collected and reported and by selecting an appropriate mix of rewards, sanctions, and technical assistance. States need to experiment, then modify their systems in light of experience.

Determining the appropriate level of accountability.

A given accountability tool, for example, a standardized test or an on-site evaluation, must be appropriately linked with the institution or person being held accountable. Aggregate or summarized data are typically too gross to be used in judging individual teachers, although they may be appropriately applied to institutions, such as districts or schools, or to all of a school's personnel.

Balancing Statewide comparability with local ownership.

Most States want data that let them compare schools and districts across the State. However, local districts and schools will use a State accountability system only if it produces information that also meets their needs and produces it on a time cycle that coincides with school-level planning. Balancing these two purposes is both a technical and a political task.

Expanding the alternatives to traditional standardized tests.

The standardized tests that constitute the core of State accountability systems are imperfect instruments. Supplements and alternatives to these tests are in the research and development stage. States and districts should support efforts to devise alternative instruments and should compensate for some of the tests' limitations by using multiple indicators of student and school performance.

Making fair comparisons.

States and districts are experimenting with ways to make comparisons that are both equitable and informative. They must also ensure that comparisons are used in responsible and productive ways. In particular, the media should be acquainted with the control factors that States or districts are using so that they can communicate comparison results clearly to the public.

Ensuring adequate capacity.

States must ensure that adequate capacity—computer systems, staff, training, and the like—is in place before implementing an expanded accountability system.

Improving the Design and Use of Performance Accountability Systems

Each State must devise an accountability system that meets its own needs and political traditions. Therefore, the Study Group did not offer prescriptions for improving these systems; rather, it presented three groups of general guidelines.

1. Recommendations for designing better accountability systems:

- Make the purpose clear.
- Include multiple indicators.
- Select indicators that are derived from State or district goals; these goals should not be limited by today's tests and measurement technology.
- Include the indicators called for by the Council of Chief State School Officers (CCSSO) among State indicators.

- Make school- and district-level accountability data public, and tailor performance reports to different audiences.
- Ensure that local districts and schools can use the accountability data, beyond State rewards or punishments.
- Establish an independent oversight body to monitor the development and implementation of the performance accountability system.

2. Recommendations for using accountability data to improve low performing schools:

- Use multiple criteria to identify low performing schools and districts.
- Provide opportunities for locally directed improvement activities before initiating State intervention; if State intervention is considered appropriate, it should proceed in well-defined stages.

3. Recommendations for using accountability data to sustain high performing or encourage significantly improving schools:

- Develop a definition of what constitutes high performance, and communicate the definition to promote consensus among policymakers, educators, parents, and students in the State. Clarify the importance of improvement over time as well as ranking among the top schools in the State.
- Use performance measures that are objective and based on multiple indicators, not on a single test score, to identify high performing or significantly improving schools.
- If State recognition of performance requires an application process, provide assistance to help districts or schools apply.
- Consider a broad range of programs and incentives.

Accountability systems are powerful tools. In their current, rather primitive state, they produce some unintended effects that can constrain instruction and shape administrative policies in ways that many find inappropriate. But they are rapidly evolving as technology improves and as the political context shifts. Attention to their current shortcomings will ensure an evolution that will harness their power to help improve American schools.

Contents

Preface iii
Foreword v
Executive Summary vii

Section 1 Introduction 1

Section 2 The Elements of a Performance Accountability System 5
What is an Indicator? 5
Criteria for Designing Indicator Systems 7
Basing Accountability Indicators on a Schooling Model 9
Summary 9

Section 3 State Performance Accountability Systems: Description and Trends 11
Types of Systems 11
Design Decisions Common to all Systems 18
Summary 30

Section 4 Uses and Consequences of Performance Accountability Systems 31
What Difference Do Accountability Systems Make? 31
State-Based Performance Accountability Systems 32
Locally Based Accountability Systems in Operation 35

Section 5 Unresolved Dilemmas 37
Balancing Oversight and Improvement 37
Determining the Appropriate Level of Accountability 38
Balancing Statewide Comparability with Local Ownership 39
Expanding the Alternatives to Traditional Standardized Tests 40
Making Fair Comparisons 41
Ensuring Adequate Capacity 43

Section 6 The Study Group's Recommendations for Improving the Design and Use of Performance Accountability Systems 45
Designing Better Accountability Systems 45
Using Accountability Data To Improve Low Performing Schools 48
Using Accountability Data To Sustain High Performing or Encourage Significantly Improving Schools 49

Afterword 51
References 53
Acknowledgments 55

Appendices A. State Accountability Study Group Members 57
B. Information Sources 60
C. Status of Council of Chief State School Officers Educational Indicators 64

Tables 1. Educational Accountability Mechanisms 6
2. Types of State Performance Accountability Systems 12
3. Types of Systems in OERI Study Group States 12
4. Decisions Common to All Types of Performance Accountability Systems 18
5. Types of Data States Collect 20
6. Background Characteristics of South Carolina's Five Comparison Groups: 1984-85 26
7. Key Dimensions of the 50 State Performance Accountability Systems 28

Figures 1. A Simple Model of the Educational System 10
2. A Complex Model of the Educational System Showing Links Among Elements 10
3. Background Characteristics Collected in Illinois 22
4. Connecticut's State-Level Indicator Reporting 25
5. An Example of California's "Floating" Bands 27
6. An Alternative to Standardized, Multiple-Choice Tests 42

Section 1

Introduction

Deeply held democratic traditions lie at the heart of public schooling in the United States. Although they are often difficult to put into practice, the notion of a common public interest in schooling and the widespread belief that educators should be accountable to the larger body politic have endured, even as the nature of schooling and students has changed. The persistence of these values stems partly from the philosophy that all government activity derives from the consent of the governed and partly from the practical belief that the more control citizens have over school policy, the more likely they are to support public education.

Concern about accountability in education mirrors a growing interest in holding all types of public and private institutions accountable for their performance. The Federal Government is currently studying ways to define and report performance in a wide variety of public services. For example, the Federal Government now regularly reports mortality statistics for hospitals and the on-time departure records of the Nation's airlines. Two related assumptions underlie the publication of this information: 1) that such data allow consumers to make more informed decisions and 2) that public reporting of performance informs service providers and motivates them to improve their product.

Accountability in public education has traditionally depended on such mechanisms for generating information about schools as financial accounts, curricula, student attendance records, test results, and the like. It is assumed that such information will serve as a resource for policymakers, concerned professionals, and the public who can use it to demand or effect improvements in schooling. The fact that the overwhelming majority of States[1] has chosen to structure accountability systems around the collection and reporting of data is evidence of a broad-based consensus.

Today's Accountability Systems

However, little consensus exists about the kinds of specific data that should be collected or about how to report and use those data to hold schools accountable. The 50 States, therefore, exhibit considerable diversity in their accountability systems.

- Some States assume primary responsibility for collecting and reporting information to the public, while others either share responsibility with local districts or delegate authority entirely to them.
- Some States assume that the unit of accountability is the individual school; others focus on local districts.
- Some States hold districts and schools accountable for student achievement in specific curriculum areas; others focus only on basic skills.
- Some States assume that accountability extends just to the point of providing the public with information that can be used to bring about changes considered appropriate. Other States believe that

accountability consists not only of reporting data about its schools, but also of linking that information to specific policies that reward good performance or seek to remedy poor performance.

In addition, accountability systems are currently in a state of flux. More than one-half the States recently reported that they are considering changes in either their student assessment or overall accountability reporting programs.

This Report

This report is designed to assist State policymakers who are interested in modifying or expanding their educational accountability systems. It resulted from the deliberations of the Office of Educational Research and Improvement (OERI) State Accountability Study Group, whose members represented 10 States (California, Colorado, Connecticut, Florida, Georgia, Illinois, Massachusetts, Minnesota, New Jersey, and South Carolina) and included governors' education aides, State legislators, legislative staff, State education agency staff, and representatives of local school districts.[2] The report describes several different approaches to accountability, focusing on the ways that States and local districts use the data generated by different accountability systems, the consequences of that use, and the ways in which data use might be enhanced to further school improvement goals. The discussion is based on information obtained in a 1987 50-State survey conducted by the Council of Chief State School Officers, the experience of the 10 States represented in the OERI Study Group, and research findings about the use and effects of State accountability systems in 4 of those 10 States.

One theme runs throughout the report: The design and use of an accountability system is hard and riddled with ambiguities. At one level, developing a system to collect data reliably and efficiently poses a host of technical challenges that are complicated by the relatively primitive state of tests and measurement technology. More important, however, accountability involves tough political choices: Who should be held accountable—local districts, individual schools, teachers, or students? How much local diversity should be considered in the way performance standards are defined? How much should States and local districts emphasize school improvement objectives relative to oversight and monitoring? If policy actions are linked to the accountability system, what is the proper mix of rewards, sanctions, and technical assistance? These are but a few of the issues that must be considered. Although this report provides guidance based on experience and research on many of the issues, it cannot supply definitive answers. The Study Group recognized that there is no *best* system. Each State must design an accountability strategy that supports its own educational goals and needs and is consistent with its own political traditions.

The structure of the report reflects the purpose of offering general guidelines and presenting a variety of approaches that States might adapt to their own needs and circumstances. Section 2 describes the basic elements of a State performance accountability system and outlines several conditions necessary for it to work effectively, regardless of its specific design or components. Section 3 describes six generic types of accountability systems,

how they operate in different States, their advantages and disadvantages, and the basic technical choices that States make in designing their own accountability systems. Section 4 shows the ways in which information generated by accountability systems is currently used by States and local districts and analyzes the consequences of its use. Section 5 summarizes major dilemmas that will continue to face policymakers in developing or modifying State accountability systems.

Following this overview and description of the analytical context for the Study Group's recommendations, the Study Group's recommendations for how these systems might more effectively inform school improvement strategies are presented in section 6.

[1] According to a 1987 survey conducted by the Council of Chief State School Officers, all but five States (Alaska, Iowa, Montana, Nebraska, and North Dakota) have provisions requiring that either the State or local districts collect data about student performance.

[2] The names and affiliations of the Study Group members are given in appendix A.

Section 2

The Elements of a Performance Accountability System

A number of mechanisms exist for promoting accountability in public education (table 1). These mechanisms serve different purposes and rely on different types of information. Although each plays an important role in holding institutions and individuals accountable, one mechanism, "performance reporting," has been selected as the focus of this study, because it helps the public learn how well schools are educating their students. Fiscal audits, school accreditation, and personnel evaluations have existed in public schools in one form or other for decades. With the advent of Federal and State categorical programs, program evaluation has also become institutionalized as an accountability mechanism. Although standardized tests and other indicators of school performance have been routinely used over the past 20 years in assessing local districts, schools, and students, the notion of combining all these data into a single system that reports overall performance and then uses it as the basis for further policy action is relatively new.

The examination of performance accountability systems begins with a description of the major components and the criteria that need to be met in constructing them.

What Is an Indicator?

Indicators, or statistics that reveal something about the health or performance of the educational system, constitute the basic building blocks of State performance accountability systems. However, not all statistics about education can function appropriately as indicators. Statistics qualify as indicators only if they serve as gauges, that is, if they tell a great deal about the entire system by reporting the condition of a few particularly significant features. For example, the number of students enrolled in schools is an important fact about the size of the educational system, but it tells little about how well the system is functioning. On the other hand, a statistic that reports the proportion of secondary students who have successfully completed advanced study in mathematics provides useful information about the level at which students are participating and achieving in that subject. This statistic provides considerable insight about the condition of the system and can be appropriately considered an indicator.

As they choose indicators to include in accountability systems, policymakers determine *who* will be held accountable, *for what*, and *to whom*. For example, if school-level scores on subject-matter tests are chosen as an indicator of student performance, policymakers are holding some entity (local schools, teachers, districts, or the State) accountable for teaching students a specific curriculum. On the other hand, if policymakers decide to use only tests of basic skills as the indicator of student performance, they are holding these individuals or agencies accountable on

Table 1.
EDUCATIONAL ACCOUNTABILITY MECHANISMS

Purpose	Use	Type of Information	Whether Publicly Reported
		Fiscal Audits	
Ensure that school funds are spent for intended purposes, using sound accounting principles	Enforce fiscal probity	Revenue and expenditure data	No, except in cases of serious violations
		School Accreditation	
Ensure that all schools meet minimum resource and management standards; may include assessment of school organization and performance	Enforce minimum standards	School records, on-site visits	No, except in cases of serious problems
		Performance Reporting	
Ascertain how well schools are providing quality educational services and producing desirable student outcomes	Compare performance over time, against standards, with other schools; reward, punish, and/or assist schools	Statistical indicator data on resources, school organization, teachers, curriculum, and student outcomes	Yes, in the expectation that policymakers, educators, and the public will act on the information
		Personnel Evaluation	
Evaluate how well individual principals and teachers are performing their jobs	Reward high performance; inform professional development and tenure decisions	Evaluations by supervisors; classroom and school observations, work portfolios	No, because of the need to meet confidentiality and due process requirements
		Program Evaluation	
Determine whether a specific policy or program is producing its intended effects	Decide whether to continue or modify program, determine future funding levels	Record data, interviews with participants, observations, various statistical data	Depends on public interest in the program

a very different standard. Similarly, if they choose to use only student performance rather than to include indicators of schooling resources and processes, the standard will be considerably narrower than it would be if all three domains were considered.

Criteria for Designing Indicator Systems

The indicators States use in their accountability systems vary considerably. However, the indicators included in any system should meet six basic criteria.[1] They should

1. Measure the central features of schooling.
2. Measure what is actually being taught or considered important for students to know.
3. Provide information that is policy-relevant.
4. Focus on the school site.
5. Allow for fair comparisons.
6. Maximize usefulness and minimize burden.

1. Measure the central features of schooling.

Performance accountability systems have traditionally focused on educational inputs (for example, per pupil spending) and outputs (student test scores). That information can tell whether educational conditions are getting better or worse, but it provides little insight into why particular trends exist or how to fix problems or replicate successes. For example, recent test data indicate that high school students have poor writing skills. But with only that information, State and local policymakers cannot answer some basic questions: Do all students have this problem? Is it due to the nature and quality of the curriculum? Or is it because teachers lack the skills to teach writing effectively?

Accountability is a blunt tool unless policymakers, educators, and the public have information that allows them to determine the likely sources of a problem and find clues about how to fix it. Consequently, accountability systems should do more than simply collect testing data. They should provide an integrated picture of the schooling environment and include data on fiscal and other resources, teachers, school organization, curriculum, and the distribution of various student outcomes (achievement, participation) across different types of students and schools.

2. Measure what is actually being taught or considered important for students to know.

Effective accountability does not require that States or local districts assess student performance on all or even most of what is taught, but it does require that accountability-related measures reflect at least a core component of what is taught. Unless the standardized tests that measure student performance include skills and content covered in the curriculum, schools will be held accountable for things they are not actually doing. Whether States choose to use basic skills or subject-matter tests, they need to consider whether what is tested matches what is taught.

On the other hand, to the extent that widespread consensus exists about the skills or knowledge students must master, such as elementary science and writing, States should also make certain that schools are held accountable for such content, whether it is included in their own curricula or not. A broad-based consensus about what students ought to know is often difficult to reach, and the definition of essential knowledge often varies considerably from community to community. However, in most States and school districts, some measure of agreement exists about the core skills and knowledge that students of a given age should possess. The size of that core may vary, but the important point is that the performance accountability system should measure what the larger community believes is important for students to know.

3. Provide information that is policy-relevant.

Accountability systems should describe educational conditions that are of concern to policymakers and can be acted upon. For example, characteristics of teacher training are policy-relevant, because these characteristics can be changed through legislation or regulations governing teacher licensing. On the other hand, data about teacher attitudes and interaction with individual students, although helpful in providing a comprehensive picture of schooling, cannot so easily be acted upon by policymakers.

4. Focus on the school site.

Both research and practice have demonstrated that the individual school building is the fundamental unit of improvement in public education. Therefore, accountability systems must collect data that can be broken down to that level. This criterion is particularly important for States with districts having schools that differ in either their student composition, organization, or leadership styles. Accountability systems that blur those distinctions, for example, by using district-wide averages, make it difficult to interpret data on educational performance or to use those data to inform school improvement strategies.

5. Allow for fair comparisons.

Given differences in student backgrounds and the resources available to schools, not all schools and students start out the same. Therefore, it is important that any comparisons be made and reported fairly. The challenge is to reflect differences among schools and students fairly without institutionalizing lower expectations for some.

6. Maximize usefulness and minimize the burden.

A critical measure of an accountability system's usefulness is whether it provides information that can be readily understood by a broad audience. Data must be organized and reported in a way that is useful not only to educators, but also to policymakers, the press, and the general public. Data

must also be timely and produced on a schedule that is compatible with policymakers' decision cycles. In addition, accountability system indicators should provide information about emerging problems such as the changing numbers and types of students in urban areas.

Minimizing the burden requires designing a system that can be implemented within given cost constraints; does not strain current levels of State and local expertise in data collection, analysis, and use; and creates a limited respondent burden. Burdens can be reduced by coordinating State and local accountability efforts to minimize the amount of instructional time devoted to testing and other data collection, while ensuring that the information needs of both States and local districts are met.

These six criteria apply to the design of any State accountability system, no matter how it is structured or what specific indicators are included. They must be met if the system is to produce accountability information that is reliable, fair, and useful for improving schools.

Basing Accountability Indicators on a Schooling Model

One final element of an accountability system is implied in the word *system*. No matter how valid and reliable any single indicator may be, it cannot be interpreted independent of other indicators that reflect the larger educational system. An obvious example is student achievement trends: Changes in student achievement over time cannot be interpreted unless there is information about other factors, such as whether the characteristics of students taking a particular test are the same or whether the curricula, student attendance, or dropout rates have changed.

Consequently, if an accountability system is to avoid reporting disjointed pieces of information, it must be based on a model of how the educational system actually operates. This model may be simple and intuitive or it can be complex. Figures 1 and 2 show a simple and a more complex model. The simple system categorizes the major elements of the educational system, while the more complex model illustrates relationships among those elements. These relationships do not constitute a model in either a strict predictive or causal sense, but they do serve as a framework, showing logical linkages among elements of the schooling system, many of which are supported by research.

Summary

To be effective, an accountability system must allow users to view performance on any single indicator within the context of the larger educational system. Relying on a model (whether simple or complex) that represents the entire schooling enterprise makes data interpretation more valid and also increases the likelihood that accountability systems will inform school improvement efforts.

---

[1] These criteria emerge both from the experience of groups that use educational indicators and from research on the design of social indicators in a variety of areas, including economic policy, education, health, and criminal justice (for example, de Neufville, 1975; MacRae, 1985; Murnane, 1987; Shavelson et al., 1987).

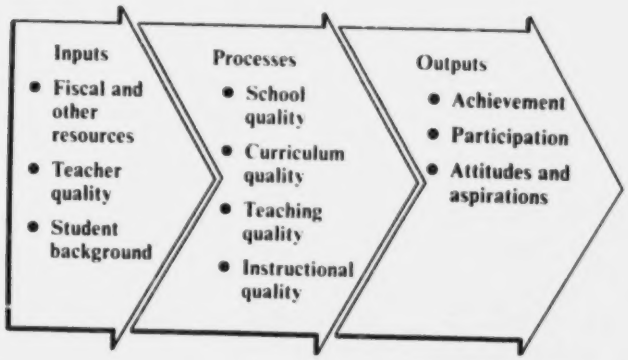
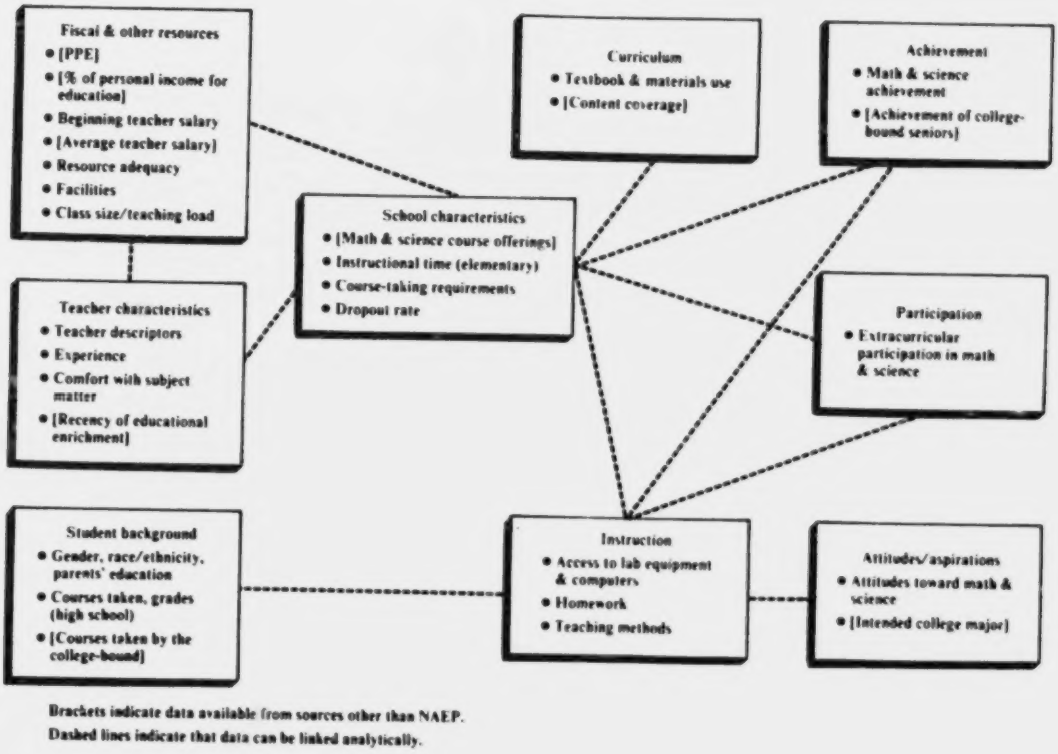


Figure 1.—A simple model of the educational system



Brackets indicate data available from sources other than NAEP.
Dashed lines indicate that data can be linked analytically.

Figure 2.—A complex model of the educational system showing links among elements
SOURCE: Reprinted with permission from pages vi and 32 of *Indicator Systems for Monitoring Mathematics and Science Education* (R-3570-NSF), August 1987, The RAND Corporation.

Section 3

State Performance Accountability Systems: Description and Trends

Types of Systems

All 50 States and the District of Columbia collect data about schooling which can be used for accountability purposes. Although all of these systems are intended to provide information to help policymakers and educators monitor the effects of policies and improve the quality of education, the systems vary considerably in their design, largely reflecting the unique historical and political circumstances of each State. Despite these differences, each system can be classified along two lines: 1) who is primarily responsible for the system and 2) whether the system is linked with other State policies.

Systems differ in whether the State or local districts have primary responsibility for deciding what data will be collected and how those data will be organized and reported. Most States have assumed primary responsibility, either as a result of specific legislation or in response to regulations developed by the State education agency. Two States require that accountability data be collected and reported, but stipulate that local districts must design their own systems and report performance. In other States, the State and local district share accountability responsibilities.

Twenty-five States with performance accountability systems link those systems with direct State-level policies. In these States, performance data about schools and districts trigger other State actions aimed at improving education. Three general types of actions have emerged among States as they have formulated these links:

1. Actions aimed at improving the quality of all schools.

2. Actions meant to reward high performing or improving schools and also to provide incentives for others.

3. Actions aimed at low performing schools. These actions are either sanctions that serve as disincentives to low performance or efforts that assist in the school improvement process.

Many States publicly recognize high performing and improving schools, and a few use high performance to trigger more substantial actions. Common mechanisms for recognition range from publishing the names of high performing schools in newspapers, to honoring principals at luncheons or ceremonies, to presenting them with plaques, banners, or other commendations. Such activities provide public acknowledgment, but they cannot be considered sustained policy actions. In contrast, seven States use

high performance to trigger more substantial actions or material rewards, such as allocation of extra resources or exemption from certain regulations or additional monitoring requirements.

In contrast to the relatively low number of States that link significant consequences to high performance, 25 States have developed policies that call for significant actions to be triggered by low performance or lack of improvement. These actions range from forms of assistance to fairly severe types of sanctions. They include requiring schools to develop plans for remedying their poor performance; providing technical assistance to develop and implement such plans; allocating additional funds to assist improvement efforts; withholding funds as a negative consequence of poor performance; declaring "bankruptcy" or "impairment" and intervening in the management of schools and districts; and consolidating poorly performing districts with others in the surrounding area. Some of these actions focus on schools, others on districts.[1]

Based on these two major dimensions, focus of control and policy links, six general types of performance accountability systems operating in the States can be identified. Their distribution is shown in table 2.

Table 2.
TYPES OF STATE PERFORMANCE ACCOUNTABILITY SYSTEMS

Type of System	Totals	Primary Responsibility		
		State	Mixed	Local
Policy Links	25	21	3	1
No Policy Links	21	14	6	1
Totals	46	35	9	2

NOTE: Five States do not report collecting performance data for accountability purposes.

Table 3 shows how these types of systems are represented among the 10 States represented in the OERI Study Group.

Table 3.
TYPES OF SYSTEMS IN OERI STUDY GROUP STATES

Type of System	Primary Responsibility		
	State	Mixed	Local
Policy Links	CT, FL, GA MA, NJ, SC	IL, CA	MN
No Policy Links	—	CO	—

The classification of States' accountability systems into types should not be interpreted to mean that the systems within a particular category are identical. States with the same general type can have quite dissimilar systems. Moreover, many States are currently expanding or modifying their systems. These changes may actually alter their system "type." For example, during the period of time the OERI Study Group met, the Colorado State Board of Education amended its accountability regulations, adding both a Statewide testing program and State reporting of local performance. These changes pushed Colorado from a State that vests primary responsibility in local districts to a State in which both the State and local districts have major responsibilities for the system. Other States in the Study Group also reported impending changes. Nevertheless, the following snapshots of representative States provide some concrete examples of various types of systems, at least at one moment in time.

State Responsibility, Direct Policy Links

Nearly one-half of the States take primary responsibility for the design and implementation of their accountability systems (by action of either the legislature or the State board of education) and also link the performance of districts or schools to other State policies. South Carolina has a highly structured system of this type, in which uniform data are collected for an integrated set of indicators on schools and districts throughout the State. Student-focused indicators include scores on norm-referenced tests of basic skills (that is, Comprehensive Test of Basic Skills) and State-developed tests of student mastery of grade-level objectives in reading, writing, and mathematics. Both types of tests are administered at several grade levels. School-focused indicators include student and teacher attendance rates and student attrition (dropout) rates.

In South Carolina, districts that fail to meet State-established criteria on a number of indicators are labeled "seriously impaired in educational quality" and are subject to State intervention. In the most extreme cases, they are subject to the withholding of funds. Schools that show greater-than-expected gains in student achievement receive monetary rewards. Moreover, the amount of the reward is increased for schools that also demonstrate gains in student and teacher attendance. Other State allocations, for example, funding for early childhood programs, are also based on the results of the assessment system (for example, the number of kindergartners in a district who fail to meet State criteria for first grade readiness).[2]

Georgia has a less complex system that assesses students' basic competencies and achievement Statewide. It does not collect an integrated set of indicators about schools and districts, but State-level policy actions are nevertheless triggered by performance on these assessments in that funds for remedial programs are allocated on the basis of students' test scores. This limited configuration is more common than the complex type of system found in South Carolina.

New Jersey has one of the most highly publicized State-controlled systems, being among the first States to link performance data with policies permitting State intervention in "educationally bankrupt" school districts. New Jersey monitors its schools on 51 indicators, including attendance,

teachers' certification status, extent of desegregation, and student performance on State tests. Policies link performance on these indicators to sanctions. Low levels of school performance set into motion an incremental process whereby the State education agency closely monitors districts' efforts to improve their schools. At each level of the monitoring process, if improvement does not occur, State action becomes increasingly prescriptive, culminating in the State's stepping in and implementing its own corrective action. In addition to linking sanctions to performance, New Jersey's policies also provide a tangible reward for high performing schools: Districts with no deficient schools receive a 5-year waiver from State monitoring, providing them considerable discretion in their programs and operations.

In contrast, Connecticut targets districts with low test scores for assistance, not sanctions. Currently, trends for remedial assistance, dropout prevention, and "priority schools" (a technical assistance program) are linked to performance on the Connecticut Mastery Test. In 1989-90, Connecticut's school finance formula will include test scores as well as a poverty index in a pupil need component. This revision will direct more general aid to those districts with low performing pupils.

State Responsibility, No Direct Policy Links

One-fourth of the States (14) take primary responsibility for the design and implementation of their accountability systems but have no other State policies linked directly to them. Nevada, for example, has recently begun to collect district-level data about teacher/pupil ratios, ethnic composition of the student body, and dropout rates, in addition to its achievement and competency testing programs. However, these data are used only for a consolidated "State of the State of Education" report to Nevada residents and do not directly trigger any other policy actions.

Arizona has a less complex system, in which the legislature requires that all students be tested in basic reading and mathematics skills each year. Individual student scores are sent home with each student on the last day of the school year, and the State publishes summary statistics. No district- or school-level analyses are made, and no predetermined policy actions are triggered by students' scores.

Mixed Responsibility, Direct Policy Links

Only three States share responsibility for the design and implementation of the educational accountability system with local districts and link their systems to other State policies.

In Illinois, the responsibility for defining, collecting, and reporting indicators is shared between the State and local levels. The State mandates that the schools develop local goals for "excellence in education" and devise procedures for assessing student achievement at grades 3, 6, 8, and 11. The State requires that, as part of these local assessments, some State designed assessment items be included so that State-level results can be reported. The State Board of Education staff reviews and approves local plans and prepares school "report cards" to provide assessment results to the public. At the local level, districts formulate local improvement goals and establish student learning objectives. Each district submits an assessment plan to the State, stipulating how it will determine whether students are meeting

these goals and objectives at the specified grades. Local school districts are responsible for distributing the State prepared school report cards to the public. They also have the option of modifying these reports, as long as all the data and explanatory text are included.

Direct policy actions based on indicator results emanate from the State and local levels. These actions include requiring schools or districts that do not meet their objectives to develop school improvement plans and providing technical assistance in the development and implementation of these plans.

California also has a system with mixed responsibility and policy links. The State collects a standard set of quality indicators from each district and reports uniformly for schools. It also mandates that each district develop a proficiency test for high-school graduation. However, California also encourages school districts to develop and report local indicators and suggests other data (for example, number of writing assignments, amount of homework) that might be collected and reported. Whether or not these local indicators are actually developed remains the responsibility of local districts. State policies are triggered by both high and low performance on multiple State indicators. The normal 3-year review cycle is extended for high performing schools, and technical assistance is provided to low performing schools. Like many other States, however, California also publicly recognizes schools showing the greatest improvement on the State achievement tests and publishes lists of the poorest performing schools.

Mixed Responsibility, No Direct Policy Links

In six States, the State and local districts share responsibility for designing and implementing the accountability system, but the system is not linked to other State-mandated policies.

The Colorado State Board of Education requires that all districts implement a process for assessing and reporting on educational achievement at the local district level, giving locals almost total responsibility over the system. State regulations mandate that the process must specify goals and objectives, activities for achieving them, mechanisms for evaluating student performance, and provisions for reporting results to residents of the district and the State Board of Education. The State also requires that the plan be developed by a group that includes at least one parent, one teacher, one school administrator, and one taxpayer, and it mandates that performance reports include information about district priorities, academic achievement, dropout rates, and revenues and expenditures. Nonetheless, the substantive nature of the goals and activities and the way in which evaluative information is reported are locally determined and vary from district to district.

As noted earlier, the Colorado State Board has recently added more direct State responsibility to the accountability system with the development of a Statewide testing plan (based on experience with an experimental Statewide testing program in 1986). However, while the new State tests are intended to be an integral part of the accountability system, it is not clear how they will link with the locally determined processes described above.

Even with this added layer of State responsibility, Colorado has chosen not to link performance results to direct policy actions. Rather, the intent of Colorado's accountability system is to hold up districts with outstanding programs as examples to others, and to offer (but not mandate) technical assistance to schools and districts with serious problems.

Oregon's system is another example of shared responsibility with no policy link. Oregon mandates an achievement assessment designed and implemented by the State, but it also calls for local districts to design and implement high school proficiency tests. Although the results of the Statewide assessment are used to identify schools needing improvement, no direct State policy actions are triggered by results from either the State or local tests.

Local Responsibility, Direct Policy Links

Only one State gives local districts primary responsibility for designing and implementing their own accountability systems, which trigger specific State policies. Current State law in Minnesota lets local districts select the learner outcomes that will be emphasized in instruction and assessed by performance measures. However, Minnesota requires that about one-third of the assessment items measure "core learner outcomes" that are specified by the State, and a State-developed item bank provides items for assessing both core outcomes and those that are locally selected. Local districts must use the items measuring the core outcomes, but they also can elect to use others that match their local priorities.

Despite its local focus, Minnesota links specific State policy actions to the results of the local assessments. The State requires that each district develop an individualized Assurance of Mastery program for any student not performing to locally set standards in mathematics and communication skills. However, the link between student performance data and State involvement in the Assurance of Mastery program is rather weak. Locals have a great deal of latitude in the design and implementation of the Assurance of Mastery requirement, and the State has few mechanisms to enforce the policy. Although the State offers districts one dollar per student for compliance with the program, some districts find this an insufficient incentive.[3]

Local Responsibility, No Direct Policy Links

Only one State, Vermont, assigns local districts responsibility for designing and implementing their accountability systems, a competency testing program. Vermont has no State policies linked to the data. However, as in other States, the absence of State policy links does not preclude local action being triggered by schools' performance on indicators.[4]

Weighing the Alternatives

States that assign the primary responsibility for the system to local districts or that have systems of shared responsibility are often States with strong traditions of local control, for example, Colorado, Illinois, Minnesota, and Vermont. The trend, however, even in these States, is to move toward greater State involvement with accountability. Assigning the State primary responsibility often capitalizes on the greater technical capacity and resources at that level. Additionally, State-controlled systems may be viewed as producing information that is more useful to interested State

policymakers (for example, governors and State legislators), because such systems can more easily produce comparable data about local schools and districts. A disadvantage of strong State control, however, is that it moves the center of accountability away from local districts and schools—the units that are most able to use information to improve education. States that are developing shared responsibility are attempting to bridge this gap. As discussed in greater detail in section 5, States must weigh these trade-offs carefully.

States that choose to link policy actions to performance data do so in order to put teeth in their accountability systems. Policymakers in these States generally believe that mandating specific consequences for various levels of performance provides incentives for improvement, tangible mechanisms for helping schools improve, or both. On the other hand, many States report that making performance data public is a powerful enough mechanism for spurring school improvement, and they have chosen not to link other specific actions with performance data. These States also make individual choices based on their experiences and their assumptions about what is needed to improve their educational systems.

Direct policy links have both advantages and disadvantages. Even policies aimed at rewarding high performance or improvement are not without detractors. Some consider ceremonies and awards to be superficial tokens that are inappropriate to the level of accomplishment schools have achieved. Others view them as cheap substitutes for more tangible rewards. Both recognition and more tangible rewards have been characterized as "dangling carrots" in front of educators; they are seen by some as policies that embody the assumption that schools will not work unless external rewards are forthcoming. Finally, in States that base rewards only on test scores, policies rewarding high performance have been criticized as being overly simplistic.

Some States, however, report positive results from recognition and rewards, and others are devising ways to keep these policies from being perceived simply as external motivation for higher test scores. Florida, Georgia, and South Carolina, for example, all note that a surge in school and district morale has accompanied recognition or rewards for schools that are doing well. Others have built into reward policies mechanisms to help keep them from being perceived as superficial or overly dependent on test scores. Massachusetts Project REACH (Rewarding Educational Achievement), for example, permits schools to apply for awards (currently $25,000 for elementary and $75,000 for secondary schools) and to base their applications on multiple criteria. Although the State requires that improved academic performance be one measure, it plans to include other indicators as well, some of which schools will select themselves as illustrative of their performance.

The remedies or sanctions a State chooses are likely to reflect its past education policies and prior efforts to improve low achieving schools. For example, Connecticut has decided to provide extra resources to assist such schools, believing that the State's past history of fiscal inequities among local districts has contributed to current achievement differences. New

Jersey and South Carolina, in contrast, levy stringent sanctions against low achieving districts and schools. Their policies stem from frustration with the limited gains that have been realized from past efforts to improve low achieving schools. In New Jersey, for example, the concern is greatest about the persistence of very low levels of performance in urban minority schools. In South Carolina, the greatest concern is with the continuing lack of improvement in small rural schools.

The advantages and disadvantages of each type of policy for improving low performing schools must be considered carefully. Special resources targeted at low performers may have the unintended consequence of providing an incentive to remain in the low achieving category. Sanctions, particularly the more severe ones, may affect the quality of the data that can be collected from schools. For example, schools may feel pressed to do whatever is necessary to look good on performance indicators, and this pressure can lead to such extreme actions as manipulation of the data reported to the State, and cheating by teachers and students during test preparation and test taking.

Design Decisions Common to All Systems

The accountability systems described above illustrate that as States construct accountability systems, they make a number of design decisions in addition to determining who has primary responsibility for the system and whether State actions will be linked to the performance data the system generates. These decisions, listed in table 4, contribute further variety, even among States with systems that are similar in terms of the locus of authority and policy links.

Table 4.
DECISIONS COMMON TO ALL TYPES OF PERFORMANCE ACCOUNTABILITY SYSTEMS

- What will the indicators be?
  - Student performance data?
  - School resources and processes?
  - Background data on students and communities?
- At what level will data be collected?
  - State? District? School? Classroom? Student?
- At what level will data be aggregated and reported?
  - State? District? School?
- What mechanisms will be used for reporting?
  - Reports to the public and media?
  - Reports to parents?
  - Make data available on request?
- Will schools be compared?
  - With all other schools?
  - With similar schools?
  - With their own past performance?

Data To Be Used as Indicators

Perhaps the most fundamental decisions States make are in 1) the choice of indicators that will be used to judge and understand the performance of students, schools, and districts and 2) the number of indicators that will be needed to provide the desired information. These decisions determine what information will be available to policymakers and educators.

In section 2, six basic criteria for indicators that are useful for school improvement were suggested. Few States have accountability systems whose indicators meet all these criteria. In many States, accountability rests largely on data provided by tests of students' learning outcomes. As yet, the most common decision appears to be that the core of the accountability system is students' performance on tests of basic academic knowledge and skills or minimum competencies. Twenty-nine States have both achievement and minimum competency testing requirements. Eighteen States rely on achievement assessments alone (either custom-made or commercial standardized tests). Only one State currently requires only minimum competency data on its students, and only three have no provision at all for Statewide student testing. In many States, these test scores are the only indicator data collected or reported.

A 1987 survey by the Council of Chief State School Officers (CCSSO) found that only 23 States go beyond test scores and use an integrated set of indicators.[5] These States often juxtapose data about students, school resources, and instructional conditions to provide a more comprehensive picture of schooling. California, for example, uses a combination of student and school indicators. Student indicators include scores on State-developed achievement tests, scores on Scholastic Achievement Tests (SAT) and Advanced Placement (AP) tests, and first-year grades of students continuing in the State university system. School indicators include academic course enrollments, attendance, and dropout rates.

Illinois reports a variety of student performance indicators, including percentage of students in each of the four quartiles on nationally normed tests, percentage of elementary students not promoted to next grade, high school graduation rates, American College Test (ACT) scores, number and percentage of ACT test takers, and the national percentile ranks of schools' ACT composite scores. Illinois also reports a number of school-focused indicators (for example, class size; high school enrollments in math, science, English, and social studies; race and gender of teachers; average years of teaching experience) and district-focused indicators (for example, average salaries for teachers and administrators and per pupil expenditures).

Another 23 States reported that they collect performance data but not an integrated set of indicators. These States appear to base their systems nearly exclusively on student test data, although most of them also collect other types of data through State accreditation, auditing, or teacher licensing procedures. However, these other data collection processes typically operate independently of performance-based accountability systems. Table 5 shows the frequency with which the States in the CCSSO survey collect various types of data.

Table 5.
TYPES OF DATA STATES COLLECT

Type of Data	No. of States
Student-focused data:	
Test Scores	
Achievement	45
Competency	30
Attendance	42
Completion of School	44
Postsecondary status	27
School- and teacher-focused data (e.g., expenditures, teacher attendance, teacher credential status)	47

The advantages of using multiple indicators are numerous. First, a single outcome measure provides only limited information on which to judge performance. For example, a State that uses only test scores may draw unwarranted conclusions about schools' effectiveness. Schools where large percentages of students become discouraged and drop out may produce high test scores, since the dropouts are often among those who score lowest on tests. Adding other outcome indicators, for example, the school's retention rate, and juxtaposing them with test scores provide a more complete picture of how well a school is doing. Other types of indicators, such as enrollments in advanced courses, can also provide valuable information with which to evaluate the school, for example, the access students have to rigorous academic content and learning experiences. Second, if States collect outcome data only to assess performance, they provide policymakers with little understanding about the schooling conditions under which those outcomes were produced, and they offer little guidance about potential targets for improvement policies.

While multiple indicators provide many advantages, there are also disadvantages in collecting too many indicators. The expense and burden of data collection imposed on educators can become great, and the system itself can become unwieldy. States must therefore attempt to select a parsimonious set of indicators that provides enough information to be valid and useful for oversight and improvement.

Inclusion of Background Data

States also must decide whether to collect data about conditions outside the school, so that school performance indicators can be placed in their broader context. This usually means collecting background information about students and their communities, for example, data about student mobility, socioeconomic status, and community wealth. Many State policymakers believe that these data indicate the difficulty that schools and districts face as they educate their particular student populations. For example, schools with high student mobility rates are expected to have greater difficulties carrying out long-term or sequential learning programs. It is not fair to hold these schools entirely responsible for the learning progress of students who may have attended many different schools over the course of a year. Other data, such as the financial resources available

locally, are sometimes thought to be useful for understanding the constraints schools face in the materials they can purchase or in the instructional services they can provide.

Forty-five States in the CCSSO survey collect background data, but only 21 report that they use those data as contextual information to help policymakers and educators interpret performance indicators. South Carolina reports indicator data in the context of comparison bands reflecting the percentage of students eligible for free lunches at the school, locally generated financing, and the percentage of first grade students scoring at or above the State standard on a kindergarten readiness test. More typically, Illinois reports achievement indicator data juxtaposed with data on student characteristics, for example, school enrollment by racial/ethnic categories, percentage of low-income students, percentage of limited-English-proficient students, attendance rates, mobility rates, and percentages of students in academic, general, and vocational tracks. The background characteristics reported in Illinois are shown in figure 3.

Collecting and reporting relevant background data, however, is not easy. The proxies typically used to assess students' economic status are imprecise. Self-reported, student data are fraught with inaccuracies, as are principals' judgments about family income and parental occupation. Data about the proportions of students eligible for free lunch or who qualify for Federal compensatory education assistance provide only gross estimates of the percentages of students who are poor and are unreliable as indicators of economic status. Finally, when test data are set in the context of information about students' backgrounds, States must consider the possible unintended consequence of reinforcing pernicious stereotypes about disadvantaged racial and economic groups.

Level of Data Collection

An important technical decision that States must make concerns the level at which data will be collected. This decision is critical because it influences the levels at which data can be organized and reported, and in some cases, it influences the accuracy of the data.

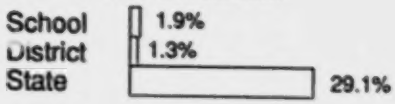
If, for example, States want schools to use performance data to assess the learning of *individual students*, they must test every student in the State, at least in certain grade levels, on all learning objectives of concern. But if States want to use student test data to assess the performance of *classrooms, schools, districts, or the State as a whole*, they can use sampling strategies. In California, for example, the school is considered the primary unit of accountability. Consequently, State tests are not expected to either diagnose or assess the learning of individual students, but rather to measure the performance of schools in helping their students learn. Therefore, California uses a matrix sampling plan, where each student answers only a subset of test items; together, students' scores reflect the performance of the school on all of the test items. Other States that want only school-, district-, or State-level scores simply collect data from representative samples of students. Of course, student level data can be organized to describe other levels. For example, Florida uses data about the universe of its students to create reports at the student, classroom, school, district, region, and State levels.

Student Characteristics

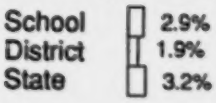
Enrollment (Racial/Ethnic Background and Total)

	White	Black	Hispanic	Asian	Native American	Total Number
School	98.7%	.0%	.6%	.6%	.0%	154
Subdist.						
District	97.4%	.0%	1.3%	1.3%	.0%	462
State	67.1%	22.2%	6.3%	2.3%	.1%	1,797,552

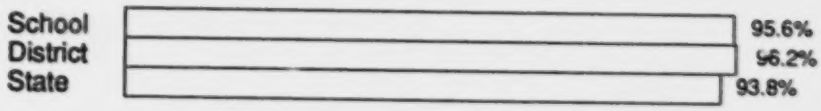
Low-Income Enrollment



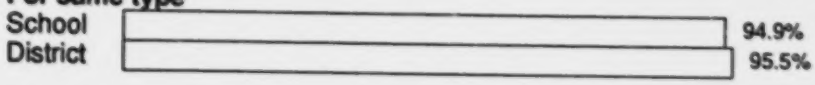
Limited-English Proficient Enrollment



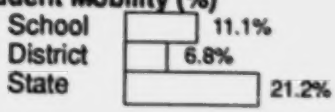
Student Attendance Rate



For same type



Student Mobility (%)



% of H.S. Senior Class who are:

College Preparatory*		General Education		Vocational Education	
School	%	School	%	School	%
District	%	District	%	District	%
State	%	State	%	State	%

*The standard by which seniors are judged to be prepared for college differs from high school to high school around the state.

Figure 3.—Background characteristics collected by Illinois
SOURCE: Illinois State Board of Education (1987).

An important rule of thumb governs sampling strategies: Data from samples can be aggregated up but not down. That is, data sampled at the school level can usually be used to describe districts and States as well, but data sampled at the district level will not provide adequate information about individual schools and certainly not about individual students.

These same considerations apply to data collection about classrooms, schools, and districts. States that collect data about school resources and processes from all schools can organize these data to provide information about all districts and about the State as a whole. Data, for example, course enrollments, that are collected about a unit from the unit itself, such as a school, may also be more accurate, since those supplying the information may be in the best position to have the facts at hand.

The current trend is to collect test data about all students (this is done in 40 States in the CCSSO report). Of the States that sample students, some collect data that provide information at the school level, some at the district level, and others at the State level. States must consider the advantages and disadvantages of various data collection strategies. If student-level indicators are collected from every student, individual teachers and schools may find them useful for improving school programs and classroom instruction if the measures are seen by locals as relevant to their own goals (a topic discussed later, in section 5). However, the burden of such data collection can be great. On the other hand, if school- or district-level data seem to be useful for oversight and improvement, then less burdensome sampling strategies can be used.

Strategies for Reporting Data

To a great extent, the way States report data is closely linked to the way the data are collected. Although States may not choose to release all possible combinations of performance data, they must decide whether and how to report indicators. Typically, student test data have been reported to school administrators, teachers, and parents, because these are the people who are intimately involved in the learning process. However, with the increasing use of these data for accountability purposes, indicators (including student test performance data) are increasingly seen as needed by the public and policymakers. Thus, decisions about the technical issues of collecting and organizing data for reporting are closely related to the more political issue of *who should be held accountable.*

Forty States, including all of the States in the OERI Study Group, make the performance data they collect, at least at the State level, public. There is consensus that test scores of individual students should not be reported to the general public, but many States provide these data to parents. However, there is considerable variety in the types of data that are reported to the public and in how those data are reported.

The national trend is to consider individual schools the primary unit of accountability. Twenty-five States report school-level summaries of test data to the public. Twelve States report district averages, but not school-level results. Only six States report State averages without district or school-level results. Considerable variety exists in *whether, how,* and *by whom* indicators other than test scores are reported. Several States

represented in the OERI study group report a number of indicators to the public in addition to test scores. California's array of indicators for each school was noted earlier. South Carolina reports student and teacher attendance and student attrition rates for each school. Connecticut reports percentage of twelfth grade students entering college-bound courses, pupil/teacher ratios, teachers' salaries, per pupil expenditures, and a host of other fiscal data about school districts. Connecticut also reports indicators at the State level (figure 4).

Among the States in which local districts determine the content of reports beyond test scores, Minnesota suggests that locals report local improvements at the district level, while Georgia requires districts to report the costs of providing schooling locally.

A final way in which States differ is in *how public* they make their accountability data. Some States report school- and district-level data directly to the media throughout the State; others leave individual school data reporting to local districts. In this latter group, some districts report school-level data to the media, and others send school reports to parents. Still other States keep a low profile, making their performance data publicly available only on request.

Each of these strategies for reporting data has consequences for schools. Most States' experience has been with the public reporting of standardized test scores. Many States have found the public to be highly interested in these scores, and not always for reasons relating to educational improvement. For example, in many States, real estate agents have found published test scores useful in marketing homes in neighborhoods with high scoring schools.

Generally, however, States report that highly visible school-by-school score reporting has triggered considerable public and school district pressure for improving test scores. Some States also report extremely low morale in those schools where low performance is made highly visible. In choosing reporting strategies, States must decide whether they desire these consequences and, if so, how they might be used as a positive impetus for improvement. Since schools will make considerable efforts to look good on whatever indicators are made public, decisions about which indicators to report are as significant as how they will be reported. Whatever States decide about what indicators to report and how to report them, they must also anticipate how to avoid unintended and undesirable effects.

Making Comparisons

One of the most politically troublesome and technically sophisticated decisions States make is whether and how to compare the performance of schools and districts with one another. Many States do make and report comparisons. In the CCSSO report, 23 States said that they actually report comparisons, and 16 more States said their reports enable comparisons to be made.

Although the national trend is toward comparing the performance of schools, of districts, and of the States themselves, making fair comparisons is complex. Some States are able to compare their schools with national

Indicators of Success

↑	Change is in the desired direction.
↓	Change is counter to the desired direction.
↔	There is little change.
—	Baseline data. No comparable data were available before 1985-86.
⊘	No data indicating change are available.

Status	Indicators of Change	Commentary
—	The number of handicapped students placed into competitive employment	In 1985-86, 1,805 of the 2,184 clients rehabilitated in programs supported by the Division of Rehabilitation Services were placed into competitive employment.
↑	The proportion of vocational education program completers who obtain a full-time job related to their training, pursue additional education, or enter the military	In 1985, 70.9 percent of all vocational education program completers either obtained a full-time job related to their training (24.3%); pursued additional education (43.5%); or entered the military (3.1%). This figure is up from 68.9 the previous year.
⊘	The number of handicapped students graduating high school and placed into postsecondary training	No data are available on the number of handicapped students graduating high school who are placed in postsecondary training. Data will be available for the 1986-87 school year.
—	The number of handicapped students in vocational education programs	19,457 handicapped students, constituting 12 percent of the total, were enrolled in vocational programs in the local public school districts in 1985-86. These data reestablish a baseline for this indicator, as the prior data were found to be significantly underreported.
↑	Percentage of public high school students enrolled in the vocational-technical schools	The percentage of public high school students enrolled in the vocational-technical schools in 1985 was 7.8. This was 0.1 percentage points above the prior year and 2.2 percentage points above the level of five years ago.
↔	The average score on the general educational development (GED) tests	In 1985, the average scores on the writing skills, social studies, reading, and mathematics sections of the GED test declined while the average increased slightly on the science section. The five components of the GED totaled 245.1 in 1985 compared to 247.3 in 1984.

Figure 4.—Connecticut's State-level indicator reporting
SOURCE: Connecticut State Department of Education (1986)

norms because they use commercially produced standardized tests. Other methods used for comparisons include ranking schools within the State and computing percentiles into which schools or districts fall. More complicated approaches have been designed to make comparisons fair by creating an "even playing field" for all schools. These methods include judging performance in terms of progress over past achievements, comparing schools and districts with similar characteristics, and predicting "expected" scores from a combination of past achievement and student demographic characteristics.

Several different methods are represented in the OERI Study Group States. One dimension of Florida's testing program is a norm-referenced test. For this program, schools' percentile placements are computed and individual schools are compared to other schools that are similar in socioeconomic status and prior achievement. The average performance in the State is then compared to national norms.

California has devised a method of "floating" comparison bands, where a school's test performance is compared with that of 80 schools that rank immediately above the school and with 80 that rank immediately below it on several background characteristics (figure 5).

California also reports scores in the context of schools' past scores, as does Massachusetts. Massachusetts is among a growing number of States that is developing regression equations that predict expected test scores from both past performance and school characteristics. South Carolina compares schools in a number of different ways, including past performance, the use of clusters of similar schools as comparison groups on achievement indicators, and percentile rankings within the State on student and teacher attendance, dropout rates, and test scores (table 6).

Table 6.
BACKGROUND CHARACTERISTICS OF SOUTH CAROLINA'S FIVE COMPARISON GROUPS: 1984-85

Group	Median Percent Free Lunch	Median Percent At or Above Standard CSAB	Median Teacher Education Beyond Bachelor's Degree (years)	Median Expense Over Required EPA Funding ($/pupil)
One	76.2	66.1	0.9	124
Two	53.9	66.9	1.0	144
Three	37.4	74.5	1.1	170
Four	27.1	75.4	1.2	179
Five	15.1	78.5	1.3	246

NOTE: Characteristics were weighted according to their relationships with test scores by using multiple regression for each grade and the subtests of reading and mathematics. This resulted in predicted scores which were translated to standardized values (Z-scores) and then averaged over grades for each year. In this way, schools were grouped on predicted, not actual, performance. These average Z-values were used to determine the group placement of each school.
SOURCE: South Carolina Department of Education (1987).

Performance Report for California Schools

C 12 NORCAL
D 34567 MIDWAY UNIFIED
S 8901234 EPIC SENIOR HIGH
1986-87 Grade 12 Enrollment = 407

Standard Score Display

Course Enrollments 1986-87	1986-87 School Standard Score	Comparison Group Standard Score Range
Mathematics		
3 or more years	488	352-608
Advanced mathematics	499	359-576
English		
4 or more years	652	368-615
Science		
3 or more years	592	349-624
Chemistry	616	337-596
Physics	590	351-614
Advanced science	423	347-615
History/Social Science		
4 or more years	651	344-636
Foreign Language		
3 or more years	522	352-553
Fine Arts		
1 year art/music/drama/dance	473	345-617
Units required for graduation	627	369-627
University of California Requirements		
Enrollments in a-f courses	424	336-592
Graduates completing a-f courses	439	358-573

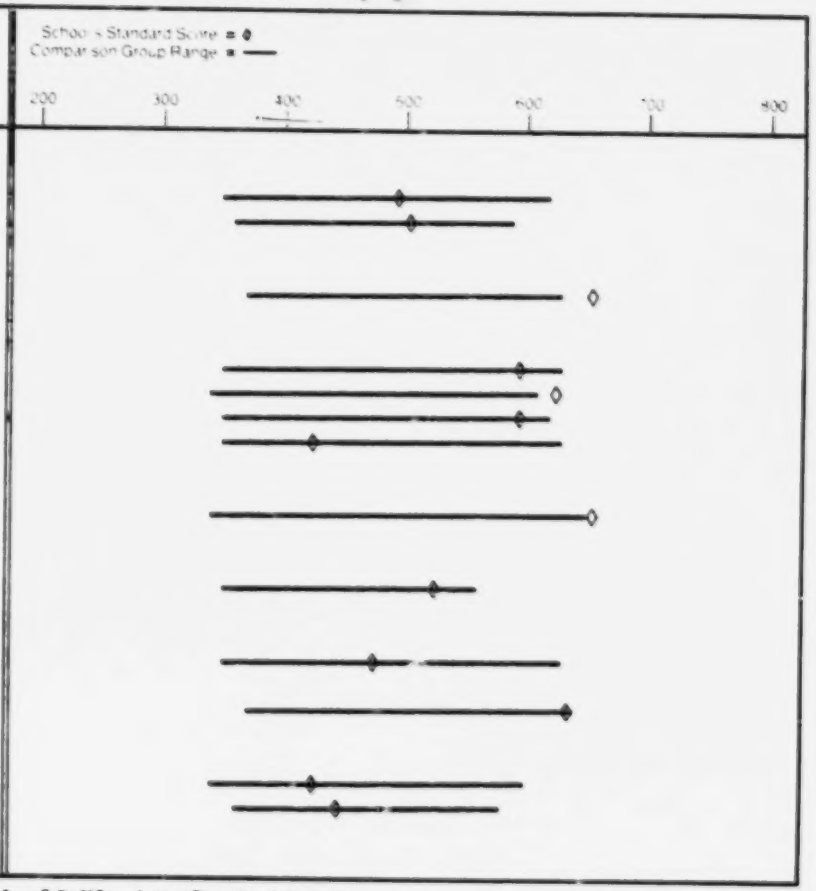


Figure 5.—An example of California's "floating" bands
SOURCE: California State Department of Education (1987)

5

ZOW

40

40

Table 7.
KEY DIMENSIONS OF THE 50 STATE PERFORMANCE ACCOUNTABILITY SYSTEMS

State	Level	Indicator System	Test Type	Public Report			Compare	Context	Policy Links
				School	District	State			
Alabama	State	No	Both	Yes	Yes	Yes	Yes	No	No
Alaska	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Arizona	State	No	Achievement	Yes	Yes	Yes	Yes	No	No
Arkansas	State	No	Both	No	Yes	Yes	Yes	Yes	Yes
California	Mixed	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Colorado	Mixed	Yes	Achievement	No	Yes	Yes	Yes	Yes	No
Connecticut	State	Yes	Achievement	No	Yes	Yes	Yes	Yes	Yes
Delaware	State	No	Achievement	Yes	Yes	Yes	Yes	No	No
DC	State[1]	No	Both	Yes	Yes	Yes[1]	Yes	No	Yes
Florida	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Georgia	State	No	Both	Yes	Yes	Yes	Yes	Yes	Yes
Hawaii	State[1]	Yes	Both	Yes	Yes	Yes[1]	Yes	Yes	Yes
Idaho	State	No	Achievement	No	No	No	No	No	No
Illinois	Mixed	Yes	Achievement	Yes	Yes	Yes	Yes	No	Yes
Indiana	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Iowa	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Kansas	State	Yes	Achievement	No	Yes	Yes	No	Yes	No
Kentucky	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Louisiana	State	No	Both	Yes	Yes	Yes	Yes	Yes	Yes
Maine	State	No	Achievement	Yes	Yes	Yes	Yes	No	No
Maryland	State	No	Both	Yes	Yes	Yes	Yes	No	No
Massachusetts	State	No	Both	Yes	Yes	Yes	Yes	Yes	Yes
Michigan	State	No	Achievement	Yes	Yes	Yes	Yes	No	Yes
Minnesota	Local	No	Achievement	No	Yes	Yes	Yes	No	Yes
Mississippi	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Missouri	State	No	Both	No[2]	Yes	Yes	No	No	No
Montana	None	No	Achievement[3]	No	No	No	N/A	N/A	N/A
Nebraska	None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Nevada	State	Yes	Both	No	Yes	Yes	Yes	No	No
New Hampshire	State	No	Achievement	No	No	No	Yes	No	No

42

41

State	Level	Indicator System	Test Type	Public Report			Compare	Context	Policy Links
				School	District	State			
New Jersey	State	No	Both	Yes	Yes	Yes	Yes	Yes	Yes
New Mexico	Mixed	Yes	Both	No	No	Yes	Yes	Yes	Yes
New York	State	Yes	Both	Yes	Yes	Yes	Yes	No	Yes
North Carolina	State	No	Both	Yes	Yes	Yes	Yes	Yes	Yes
North Dakota	None	No	Achievement[3]	No[2]	No[2]	Yes	Yes	No	No
Ohio	Mixed	Yes	Both	Yes	Yes	Yes	No	No	No
Oklahoma	State	No	Achievement	No	Yes	Yes	Yes	No	Yes
Oregon	Mixed	Yes	Both	No	No	Yes	No	No	No
Pennsylvania	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Rhode Island	State	Yes	Achievement	Yes	Yes	Yes	Yes	No	No
South Carolina	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
South Dakota	State	No	Achievement	No	No	No	No	No	No
Tennessee	State	No	Both	No	Yes	Yes	Yes	No	No
Texas	State	Yes	Both	Yes	Yes	Yes	Yes	Yes	Yes
Utah	Mixed	Yes	Both	No	No	Yes	Yes	Yes	No
Vermont	Local	Yes	Competency	No	No	N/A	N/A	No	No
Virginia	State	No	Both	No	Yes	Yes	Yes	No	Yes
Washington	State	Yes	Achievement	No	Yes	Yes	Yes	Yes	Yes
West Virginia	Mixed	Yes	Both	No	Yes	Yes	Yes	No	No
Wisconsin	Mixed	No	Both	No	No	Yes	No	No	No
Wyoming	State	No	Achievement	No	No	Yes	No	No	No
Totals	S=35 L=2 M=9 None=5	Yes=23 No=25 N/A=3	A=18 C=1 Both=29 N/A=3	Yes=25 No=23 N/A=3	Yes=37 No=11 N/A=3	Yes=43 No=4 N/A=4	Yes=38 No=8 N/A=5	Yes=21 No=26 N/A=4	Yes=25 No=22 N/A=4

SOURCE: Council of Chief State School Officers 1987 Survey and related State documents.

1 The District of Columbia and Hawaii each operate a single system in which the State and the district are the same.
2 Missouri and North Dakota send school—(plus North Dakota district)—level data to parents but not to the press.
3 Montana and North Dakota offer local districts the option of using a State achievement test.

Each of these methods has its problems, as does the whole concept of making fair comparisons. Many of the difficulties are the same as those that arise in reporting performance data in the context of student background data (see also section 5).

Summary

Each of the design decisions described in this section shapes the content and form of a State's accountability system. As the research suggests, some decisions lead to greater system effectiveness, make implementation more feasible, and generate more support. However, the decisions that lead to these desired ends differ among States, depending on current political context and policy precedents. Consequently, no best set of design decisions can be prescribed for all States, and, not surprisingly, while there are trends, no two States have made identical decisions in all areas. Table 7 illustrates the diversity among the accountability systems of the 50 States and the District of Columbia.

It is too early to judge how well each of these systems will accomplish their dual goals of oversight and educational improvement. However, some insight can be gained from analyzing the experience of four of the Study Group States. Those data are reported in the following section.

---

1 Another class of policy actions target individual students. A growing number of States use test scores and other criteria (for example, grades, courses completed) to determine the type of high school diploma a student receives (academic or vocational) or whether a student receives a diploma or merely a certificate of attendance.

2 Despite its strong State focus, South Carolina is not without local participation in the accountability process. Within a State-mandated framework, districts and schools have considerable latitude in the process and content of the improvement plans they develop. Additionally, local districts are given responsibility for reporting school-level performance data to their communities.

3 Minnesota's accountability system is currently being reconsidered by the State legislature, and proposed changes may modify its local focus. The suggested changes include comparison of districts' scores on the "core learner outcome" items and publication of these comparisons by the State. If these changes were to be implemented, Minnesota's locally oriented accountability system would be shifted toward one in which the State and locals share major responsibility.

4 Although not a performance reporting system, Nebraska's performance-based accreditation system uses a similar, locally oriented accountability framework.

5 The CCSSO survey provided national data on accountability data collection and reporting. All 50 States and the District of Columbia responded to the survey. Data from three territories were collected but are not reported in this paper.

Section 4

Uses and Consequences of Performance Accountability Systems

The different ways States can organize their accountability systems and the basic choices they need to make in designing them have been surveyed. This section examines how local districts and schools in States with accountability systems actually use the data they produce and how the systems influence school and classroom activities.

What Difference Do Accountability Systems Make?

In designing performance accountability systems, State policymakers make certain assumptions about how the resulting data will be used for State purposes and also by local districts and schools. For example, one component of Florida's testing program requires that the individual student must perform at certain established levels or a high school diploma, and possibly grade level promotion, will be denied. Remedial instruction is to be provided whenever students are not mastering required skills. The programs also require schools, districts, and the State itself to demonstrate educational achievement, especially over time, or face possible public embarrassment through release of the data. Finally, the Florida programs provide sufficient information for educators to continuously review the curriculum content and instructional programs and adjust their programs to improve educational services.

However, the way the accountability systems are used and the effects they produce depend on the response of local educators, namely, how they choose to incorporate their requirements and the information generated by them into school practice. This process significantly influences the extent to which accountability systems actually operate as their designers expected and the extent to which they produce unintended consequences.

To answer the question of how accountability systems are influencing schools and classrooms, extensive interviews were conducted with more than 350 policymakers and practitioners in California, Florida, Georgia, and Minnesota, at the State, district, school, and classroom levels.[1] They were asked

- How State accountability systems affect school-level activities and classroom teaching.
- What kind of resource burden (time and money) the accountability systems imposed.
- How accountability results are used.

The investigation shows that accountability systems produce mixed results. They are very powerful policy tools that have significantly changed school-level planning and teaching activities. But in that power lie some potential problems. Many respondents felt that the current systems motivate

teachers and administrators to focus narrowly on the specific content of tests and create incentives to manipulate their results. However, precisely because the systems have such power to change behavior, they hold the potential for dramatically enhancing the school improvement process. As accountability systems evolve, especially if they incorporate more sophisticated measures of student achievement and stress other indicators as well as achievement, they can inform the overall schooling process.

The investigation also shows that the experience of the three States with strong State components differs from that of the State with a locally designed and operated system (Minnesota). Thus, State and local systems are treated separately in this discussion.

State-Based Performance Accountability Systems

Performance accountability systems have caused local districts and schools in California, Florida, and Georgia to change the way they plan and teach. However, these changes typically are not directed at improvement, in the broader sense of using data to pursue institutional and curricular goals more effectively. Rather, accountability systems have caused principals and teachers to change their behavior because they want their students to perform well on State tests.

Instructional Effects

The comments of several principals are illustrative:

> **I pay very close attention to the [State test] results. Test results are discussed with the faculty closely, and the results are known throughout the school. I tell teachers what is in the tests, and help them develop materials that will help benefit us. We want kids to score well, feel proud, and be proud. I want the community to think we are doing a good job.**

and

> **I review the results of these tests very carefully. I look to see if children are achieving as well as we expected . . . . I assess the need for remediation in specific skills by doing an item analysis. These data are used by junior high students and aides who tutor our students on individual items . . . . We have aides who work directly with children to remediate deficient items. These aides are directed by a teacher to work with deficient children who score low on the tests. We have four of these aides in our building.**

and

> **Those scores basically define our direction in the school. We analyze the scores and adjust our teaching accordingly—for example, if our students score low on the use of apostrophes. I just finished writing my goals for the coming school year, by which I'll be evaluated. Part of those goals are based on the school's performance this**

> **year on the Basic Skills and Functional Literacy tests. I know that we are weak in writing composition, so one of my goals for the coming year is to improve students' composition skills.**

Like principals, a majority of the teachers in the fieldwork sample are concerned about their students' test performance and are willing to change their teaching practices to improve it. The following sentiment, expressed by a middle school teacher, is typical:

> **There is more thrust to teach toward the test. We also make an effort to teach test-taking skills. I don't feel that the material on the test is so important to teach, but I feel it is necessary to teach that material so the students will do well on the test. The highest stress time is during testing week. I want my kids to achieve and be successful.**

Many of the teachers interviewed said that they spend time in drill and practice aimed at improving test scores (for example, daily instruction on specific skills that are covered in the test); teaching test-taking skills, such as how to complete an answer sheet and work under time constraints; and aligning the curriculum with the tests. Teachers express substantial ambivalence about "teaching to the test," that is, teaching skills or content that is covered on a State test in the way it appears on the test.

On the other hand, they have little or no problem with teaching test-taking skills. Florida districts have made particularly extensive efforts to assist students who perform poorly on State tests. Students who are in danger of failing the functional literacy test receive special tutoring and are often placed in smaller remedial classes. In addition to providing special services for these students, schools implement schoolwide activities to improve test scores. For example, one person said:

> **When we were deficient two years ago, we identified deficiencies by item. We spend 30 minutes a week on skills. We came up nine points two years in a row. We worked on five areas, including fractions and rounding numbers. On those we did well, and that brought us up. It's a total school effort. This school has a wide array of activities to improve scores, such as mini-lessons once a week and videotaped whole school lessons to improve scores. But are we concerned with the school reputation rather than the kids? All in all, though, the State requirements make us concentrate on things and try to improve.**

The extent to which an accountability system affects school practice differs across schools. The lowest achieving schools tend to change their practices more than higher achieving ones; and elementary schools react more than high schools, largely because of the basic skills focus of State tests.

Resource Burden

The time and money spent on taking tests, collecting the data, and reporting the results vary across districts, but generally the resources required to administer tests do not impose a significant burden. In most schools, the system is effectively organized so that teachers are bothered little by test logistics. The real resource issue is the amount of instructional time schools report spending to prepare students to take the test. Most districts have their own testing systems which run parallel to the State programs, thus considerably increasing the time students spend taking standardized tests.

With regard to reporting, school personnel perceive the need for making fair comparisons across different kinds of schools. However, attempts to provide fair comparisons are not always successful. For example, although California uses a "comparison band" (figure 6) to make test results comparable across schools, not everyone—either in the schools or in the community—has accepted the notion that a similar ranking has the same meaning across bands.

Results

Data from California, Florida, and Georgia suggest that State accountability systems are producing some effects consistent with State policymakers' intentions. Local districts and schools take the existence of such systems seriously and have changed their behavior in order to improve performance. However, that performance has been narrowly defined by the focus and scope of State tests.

Although school personnel are actively responding to these systems, many do not view them as useful or valid measures of their schools' performance. Fewer than 20 percent list test scores as the most important source of information about a district or school; indeed, half the superintendents and principals in the fieldwork sample did not even mention State test scores as among the major sources of information they use to determine how well their districts or schools are doing. Their response to these systems stems from the considerable pressure they feel about having their students do well. This is the case even where there are no explicit rewards or punishments attached to test scores, just the possibility of "looking bad" (for example, appearing on a "deficient" list or receiving adverse publicity in the media).

Local superintendents and principals recognize the need to use other indicators as well as test scores in judging a district's or school's performance. They typically use a mix of quantitative and qualitative indicators such as graduation rates, the number of students attending college, teacher attendance, classroom observations, talking with students, and overall school atmosphere. However, they perceive the State as being primarily concerned with test scores, and consequently they have channeled much of their energy into improving performance on that one indicator. This response has occurred even in the two States that report having an integrated set of indicators, which suggests that if States want locals to pay attention to other indicators (for example, course enrollments, teacher

quality, the distribution of educational outcomes across different kinds of students), they will need to emphasize these indicators more strongly, perhaps by stressing them when reporting comparisons.

The accountability systems in California, Florida, and Georgia appear to have minimized the logistical burden of data collection and testing for local schools, but the amount of instructional time used for State and local testing programs and for review in preparation remains an issue.

Locally Based Accountability Systems in Operation

Minnesota has a locally based accountability system. The results of interviews suggest that in districts in Minnesota where school officials are committed to using accountability data to inform a school improvement process, the system influences school practice more broadly than it does in States with a State-operated system. However, whether this process occurs or does not depends entirely on local priorities and capacity. Some districts use the data extensively to inform practice, while others do little more than meet State requirements to test.

Some districts collect a variety of data in addition to test scores—for example, surveys of student and parent attitudes and followups on recent graduates. The district publicly reports this information and helps individual schools use the data to guide their school improvement plans. Other districts meet State reporting requirements, but they tend not to use the data once it has been published.

Local districts also vary considerably in their responses to the State Assurance of Mastery requirements. One district in the fieldwork sample has established an elaborate accountability process that includes a districtwide high school competency test to determine whether or not students are to be granted diplomas, and a set of student outcome objectives that teachers are required to incorporate into their teaching. This latter measure requires teachers to submit weekly lesson plans and to be prepared for random classroom observations by teams of administrators who will make certain that the teaching is linked to the district's student outcome objectives. On the other hand, another district in the sample has not yet implemented any program to meet the Assurance of Mastery requirements, and yet another has designed a strategy that teachers, counselors, and principals believe will have little effect on the way low achieving students are taught.

In practice, the effects of some locally designed accountability systems on school and classroom practice may be similar to the effects of State-operated systems. For example, teachers and principals in the Minnesota district that has an elaborate accountability process may see their behavior as circumscribed by the focus and scope of tests, as do their colleagues in Florida, Georgia, and California. The difference is that the accountability standards and procedures in Minnesota have been set by the local district rather than by State policymakers. On the other hand, in those Minnesota districts that, by preference or capacity, collect and use only limited data, the accountability system has very little effect on principals' and teachers' behavior.

1 This discussion is based on interview data collected in four States of the six included in research conducted by the Center for Policy Research in Education (CPRE) between 1985 and 1987 on the implementation and effects of a variety of educational reform policies, focusing on State efforts to monitor educational progress. Field interview data were collected by researchers at RAND, Rutgers University, and the University of Wisconsin-Madison. Interviews were conducted in 14 local districts and 40 schools in these four States. Approximately 120 State-level respondents were interviewed in the spring of 1986; they included governors' aides, State legislators and their staffs, State Board of Education members, State Department of Education officials, and interest group representatives. In each State, staff responsible for the State accountability program were interviewed. Interviews in local districts and schools were conducted in February–March and May–June 1987. Respondents included: superintendents, school board members, district personnel administrators, the district testing director, principals, counselors, and some 120 classroom teachers.

57

49

Section 5

Unresolved Dilemmas

The task of designing an effective accountability system and using its results appropriately is hard and riddled with ambiguities. System development involves a series of technical decisions (for example, which indicators to include and the level at which to collect data). These decisions have far-ranging political implications. For example, how much support the system will enjoy among various constituencies, the funding available to support it, and the balance between State and local concerns must be taken into account. In its operations, an accountability system may generate consequences different from those intended by its designers.

Many of these technical and political issues can never be completely resolved, but careful attention to the major problems policymakers face in developing or modifying such a system will increase its overall effectiveness. This section examines six unresolved dilemmas. The choices they pose involve core political values such as the amount of control higher levels of government should exert over school and classroom activities and the appropriate balance between "carrot" and "stick" approaches to accountability. Since consensus on these issues is rarely easy to achieve, the problems discussed here constitute some of the toughest challenges that system advocates and designers of accountability systems face.

Balancing Oversight and Improvement

Inherent in the concept of accountability is the notion of oversight—monitoring to ensure that public monies are spent in ways that produce acceptable levels of educational performance. But judging overall quality can play only a limited role in helping schools if it cannot also remedy identified problems and enhance organizational effectiveness. Most State accountability systems, at least implicitly, have oversight and improvement as their dual purposes, but the improvement function has often been a poor second. Although States and local districts have espoused the notion that accountability systems should both assess performance and inform the improvement of practice, they have made little use of accountability data beyond efforts to increase test scores. Recent initiatives that seek to strengthen curriculum frameworks and textbooks and link standardized tests to them, and that use accountability data to allocate resources and improve instruction, are beginning to equalize the two functions.

The challenge is to structure an accountability system that provides information that is useful for improvement and still makes clear that oversight is taken seriously. This can be achieved through 1) the types of data that are collected, reported, and used and 2) the mix of rewards, sanctions, and technical assistance linked to the accountability system.

Oversight and improvement will be more evenly balanced if information in addition to test scores is emphasized (for example, the distribution of course-taking across different types of students or students' postsecondary attainment). As discussed in section 4, States send strong signals with their decisions about what data to report and how to use those data. Achievement data will always constitute the core of any accountability system. However, making other information available and identifying it as valuable greatly increases the probability that accountability systems can inform school improvement in the broadest sense.

One way to supplement achievement data is to integrate existing data more effectively. States typically use such mechanisms as accreditation, fiscal audits, and teacher licensure commissions to collect a variety of data beyond what is available from their accountability systems. However, they rarely integrate these additional data into a single system that could inform the entire schooling process. This need for integration also extends to units within State departments of education and local district offices. For example, there is often little coordination among testing and evaluation bureaus, offices responsible for technical assistance to categorical programs or for general school improvement, and staff development units.

Maintaining an appropriate balance between oversight and improvement also requires policy decisions about the proper mix among rewards, sanctions, and technical assistance strategies. Since States have only recently linked significant policy actions to their accountability systems, it is too early even to suggest what the optimal mix might be for different types of States. There is virtually no information on the effects of using accountability system results to reward high achieving schools or to provide special resources for low achieving schools. It is not known whether awards are primarily symbolic or whether they purchase capacity and motivation that are not possible without them. Neither is it known what the optimal award level for sustaining high performance without causing serious inequities across districts and schools is. Past research and experience clearly demonstrate the need for focused technical assistance, yet the relative effectiveness of different forms of assistance, when tied to accountability data, is still largely unknown. Similarly, States have little experience with more severe forms of sanctions, such as intervention in local district management. It is known, however, that sanctions are inherent in accountability systems, but no one wants to look bad, and everyone will work hard not to do so, even without explicit sanctions.

Given these uncertainties, States are likely to experience a period of experimentation where they will try new approaches, which they will then need to modify in light of subsequent research and experience.

Determining the Appropriate Level of Accountability

Although State governments have constitutional authority to provide public education, they share responsibility for implementing this authority with local districts, individual schools, principals, teachers, and even students themselves. The division of responsibility among these institutions and people is sometimes quite clear. For example, responsibility for adequate financing rests with States and local districts because they are the only

institutions with taxing powers. In other areas, the division is not as apparent. For example, States that now grant high school diplomas on the basis of competency test results are assuming that students are ultimately responsible for meeting certain standards. However, in requiring districts and schools to provide remedial assistance, they are also expecting them to assume some responsibility for student performance. The tough question is how much responsibility each player should bear, particularly when negative accountability results can profoundly shape students' lives.

The question of who should be held responsible for what outcomes becomes a major issue in matching an accountability instrument (for example, a standardized achievement test or on-site personnel evaluation) with the institution or person being held accountable. The problem has recently crystallized around the issue of linking teacher and student performance and whether one can use indicator data such as scores on standardized tests to determine appropriate compensation for teachers.

Matching teacher and student performance is very difficult. Research shows that variability in educational practice and student outcomes is often greater within the same school than across multiple schools. Therefore, aggregate data or data that can be summarized statistically are typically too gross to be used in evaluating individual teachers. Student test scores are not valid proxies for teacher performance, given all the intervening factors, for example, whether the test measures what teachers teach, whether teachers are assigned subjects for which they are trained, what knowledge and skills students bring with them to the classroom, and whether a particular student is even in the teacher's classroom for the full year. The data collected must be tailored to what individual teachers actually do (for example, how they teach), which usually requires a variety of different kinds of information, including classroom observations.

On the other hand, indicator data are appropriate for holding accountable a State government, a local district, an individual school, or the principal and teachers representing a school as a group. As pressure for greater accountability in public education mounts, policymakers and the public must understand the purpose and limits of different accountability instruments. Those instruments can serve as fair and reliable measures of performance only if they are appropriately matched with the unit being held accountable.

Balancing Statewide Comparability with Local Ownership

The issue of balancing Statewide comparability with local ownership goes straight to the heart of maximizing the usefulness and minimizing the burden of a State accountability system. For a variety of reasons, most States prefer some comparability of data. The people are interested not just in the schools in their own local communities, but in the State as a whole. Taxes support all those schools, and concerns about the link between education and economic development extend beyond the local community. Furthermore, comparable data permit State policymakers to judge the similarity of educational opportunities and quality across different districts and types of students. This comparison is particularly important in States with heterogeneous or highly mobile populations.

At the same time, local districts and schools will use a State accountability system for their own purposes only if the type of information needed Statewide is also consistent with local needs and priorities. Several decades of research on the implementation of "top-down" policies has shown the importance of local ownership in influencing whether or not a policy produces its intended effects. For example, if State accountability data are to be used to inform local improvement efforts, the system must not only contain the type of information teachers need to assist individual students, but it must also produce those data on a time cycle that coincides with school-level planning requirements (for example, in assigning students and allocating resources).

At one level, balancing these two purposes is a technical challenge, that is, to obtain comparable data that are sufficiently detailed cost-efficiently. At another level, however, it is a political task, that is, to reach a consensus between State and local policymakers and between policymakers and practitioners about the purpose and priorities of an accountability system that can serve diverse constituencies. This consensus does not yet exist in most States. As indicated earlier, the need for State-level accountability and comparable data dominates systems in which the State has primary responsibility, and even in those with locally operated systems, pressure is growing for greater comparability across districts.

One concrete action that States can take in the short term is to coordinate State and local testing and other data collection activities. California is now conducting a pilot study of a comprehensive testing program that would combine the State system, which produces only school-level data, with local testing systems designed to obtain data on all students. Another strategy for coordinating State and local data collection is Minnesota's use of a test item bank that includes items measuring core learner outcomes that all districts must use, as well as thousands of other validated items that local districts and individual teachers can elect to use for measuring local curriculum results.

States might also consider designing accountability systems that allow some measures to be defined locally or others to be included or not at local discretion. If these varied local data were collected within the framework of a common model of schooling, they could be linked to data that were comparable across the State, and their relationships reported in a consistent, if not always uniform, manner.

Expanding the Alternatives to Traditional Standardized Tests

Large-scale testing programs form the core of State accountability systems and, often drive the local response. Yet, they are at best imperfect instruments for measuring academic performance. Questions about the current status and use of standardized tests are being raised increasingly by researchers, the media, and citizen groups (for example, Congressional Budget Office, 1987; Savage, 1984; Friends for Education, 1987). The most common criticisms include the inability of existing tests to measure the full range of achievement, particularly higher order skills; the limited overlap between the substance of standardized tests and the skills and content stressed in widely used textbooks; a general narrowing of the curriculum as

a result of the tests' emphasis on basic skills; and the confusing messages transmitted to policymakers and the public as a result of the long intervals between revising norms of standardized tests.

One way to redress the limits of standardized tests has already been suggested: Accountability systems should rely on multiple indicators of student and school performance. Dropout rates, the number and type of students enrolled in different courses, and student attitudes and aspirations are all examples of indicators that can be used to supplement standardized test scores to measure the health of the education system.

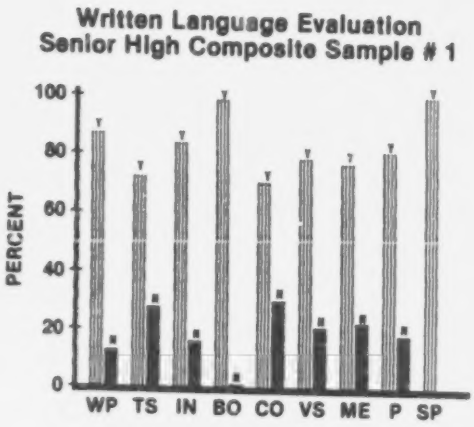
A number of States are also experimenting with enhancements of traditional standardized tests. For example, California, Connecticut, and Minnesota districts (figure 6) are now using writing samples to test student writing skills. In assessing students' abilities to apply rules of language usage and to present logical arguments, these tests can provide more valid data about thinking and reasoning skills than can a multiple-choice test. However, the use of writing samples on a widespread basis would probably require States to spend more to ensure reliable scoring.

Other supplements or alternatives to standardized testing are now in the research and development stage. The Educational Testing Service (ETS), at the request of the National Science Foundation (NSF), is designing a test of science achievement which presents students with a set of laboratory apparatus and asks them to perform a scientific task. Although such an approach may measure aspects of scientific achievement not tapped previously, it is difficult to administer and score, and it is also costly. Consequently, this technology is not yet cost-efficient for large-scale data collection. Other alternatives include using computers to simulate tasks that require the application of higher order skills, and using teacher judgments, collected through a structured survey instrument, to supplement standardized test scores (for example, see Calfee, 1988).

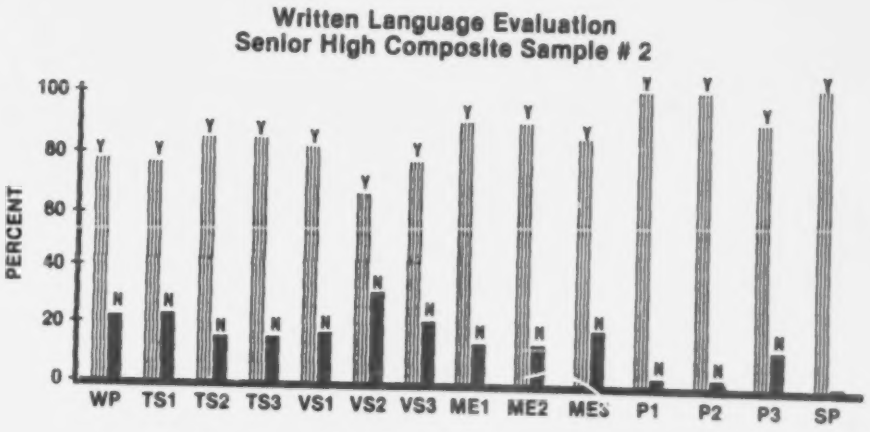
States and local school districts are not typically in a position to provide total funding for the research and development efforts necessary to develop alternatives to standardized tests, but they can motivate others to share that responsibility. Just as States have pressured textbook publishers to improve the quality of their products, they can demand that commercial test publishers develop tests that measure higher order skills more validly, and that they revise the norms of their tests more frequently. In addition, States and local districts can encourage Federal agencies responsible for research and development, such as the Department of Education and the NSF, to devote more resources to designing enhancements and supplements to standardized tests. They can also make certain that the research and development efforts of these agencies are sensitive to the constraints of time, cost, and expertise within which States and local districts must operate their accountability systems.

Making Fair Comparisons

States and local districts are struggling with the problem of how to compare districts and schools fairly. No single approach has emerged as the optimal strategy. Most people agree that it is only fair to take into consideration

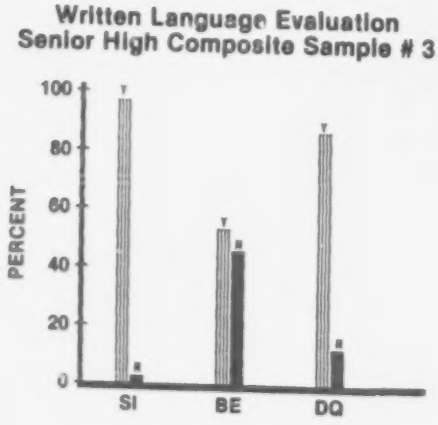


WP = Writing Process (Pre-writing, writing, revision)
TS = Thesis statement or main idea
IN = Introduction
BO = Body
CO = Conclusion
VS = Variety of sentence patterns
ME = Endmarks, internal punctuation, and capitalization (Mechanics)
P = Pronoun case and number
SP = Spelling



WP = Writing Process (Pre-writing, writing, revision)
TS1 = Topic sentence, paragraph 1
TS2 = Topic sentence, paragraph 2
TS3 = Topic sentence, paragraph 3
VS1 = Variety of sentence patterns, paragraph 1
VS2 = Variety of sentence patterns, paragraph 2
VS3 = Variety of sentence patterns, paragraph 3
ME1 = Endmarks, internal punctuation, and capitalization (Mechanics), paragraph 1
ME2 = Endmarks, internal punctuation, and capitalization (Mechanics), paragraph 2
ME3 = Endmarks, internal punctuation, and capitalization (Mechanics), paragraph 3
P1 = Pronoun case and number, paragraph 1
P2 = Pronoun case and number, paragraph 2
P3 = Pronoun case and number, paragraph 3
SP = Spelling (0% to 20% errors)

BAR KEY: Y = Yes N = No



SI = List four sources of information on assigned topic
BE = Gather sufficient information to make bibliographical entry
DQ = Include direct quotation from one of the sources

BAR KEY: Y = Yes N = No

TESTING PROCEDURE

In April 1987 all 12th grade students were given one of three writing assessment tasks to perform. Six of seven seniors completed Samples 1 or 2, and one out of seven seniors participated in Sample 3, a research exercise in the media center.

The readers evaluated 100 student writings (50—Tech, 50—Apollo) of Sample 1 and 100 student writings (50—Tech, 50 Apollo) of Sample 2. A total of 68 Sample 3 student writings (35—Tech, 33 Apollo) were also evaluated. Student writings were selected randomly but in the proper ratio from the three levels of courses offered to students. Using specific criteria, the readers tabulated the results of the writing samples.

Sample 1 was designed to assess seniors' abilities to use the writing process to compose an essay with thesis (main idea) statement, introduction, body, and conclusion. Readers also evaluated students' abilities to generate a variety of sentence patterns and use proper writing mechanics.

Sample 2 was designed to assess the abilities of seniors to utilize the writing process to compose three paragraphs: narrative, descriptive, and persuasive. Readers also evaluated students' abilities to generate a variety of sentence patterns and use proper writing mechanics.

Sample 3 was designed to assess how successfully seniors could locate sources of information about an assigned topic in the media center. The task also included gathering the information necessary to make a direct quotation from a source and make a bibliography entry.

COMPOSITE OF RESULTS

The composite charts on page 3 and at right are a summary of the writing sample results. The learner outcome number and parallel test questions followed by percentage results are listed in the first three columns.

Test results of 80% accuracy met the minimum standard; those within the range of 80-89% accuracy met and exceeded expectations; and those over 90% exceeded expectations. Test results below 80% accuracy did not meet expectations.

Figure 6.—An alternative to standardized, multiple-choice tests
SOURCE: District 742 Community Schools, St. Cloud, MN (1987)

55

significant differences in resource levels and in the students being taught at different districts and schools. However, translating that principle into practice is difficult. For example, the use of predicted scores can institutionalize low expectations for some schools; comparison bands may make some schools feel they are "the best of the worst." Some States and districts also report that because the people tend to think in absolutes, it is confusing to grasp the controls for wealth and student characteristics, especially when schools viewed by the community as low performing are designated worthy of merit. Some States and districts have experienced additional problems when they have given extra funds to schools that enroll more difficult to educate students, because, in some cases, the additional funding has proven a disincentive to improvement. On the other hand, failing to control for differences in such factors as student mobility or the number of non-English-speaking students seems to violate basic notions of fairness and creates disincentives for serving those in greatest need.

States continue to seek more valid approaches for comparing districts and schools. For example, South Carolina is using multiple methods to deal with the dilemmas that comparisons present. To relate their performance to that of other schools, schools in South Carolina can compare themselves with a group of schools having similar characteristics (table 5). However, in making school incentive awards, the State compares schools with their own performance over time, using expected performance levels based on a matched longitudinal analysis of the test scores of students in that school. California has also sought to improve its method for comparing schools. Its original approach was to group schools in five comparison bands based on student characteristics, but this put schools at the top or bottom of a band at a disadvantage. By using floating comparison bands, the State gives each school the opportunity to compare its performance with that of 80 more advantaged and 80 less advantaged schools.

States and districts continue to experiment with improved technical methods for comparisons. However, the comparisons must be used in responsible and productive ways, especially in the media's reporting of comparisons. In some States where controls for school context are used by the State education department in its reporting, the largest newspapers in the State continue to report only raw test scores; in other States, however, the newspapers regularly place schools in context when reporting their test scores. If a State or district decides to build control factors into its comparisons, the news media must be acquainted with those factors and with how to clearly communicate their meaning to the public.

Ensuring Adequate Capacity

The expansion and heightened visibility of accountability systems require new roles for State education agencies, local districts, and individual schools. However, many State governments pay little attention to the resources and capacity these new functions require of each organization. For example, a State operated accountability system designed to inform school improvement activities assumes that State departments of education can move beyond their traditional role of testing students and simply "counting things" (numbers and types of students, program offerings, library size) to assessing the entire schooling process and assisting local

districts in the use of their data. Yet this new charge comes at a time when many State agencies are implementing a variety of other reforms. Some have faced budget and staff reductions because of decreased Federal and State funding, and many are constrained by the staffing requirements of categorical programs.

Similarly, accountability systems place new responsibilities on local districts and schools. Many large urban districts and affluent suburban ones have traditionally had the capacity to collect and use data in ways designed to aid school improvement. However, in most States, local capacity levels vary significantly. These differences are particularly marked in States with small rural districts or especially poor ones. States need to consider whether adequate capacity—whether it be computer systems, additional staff, or training—is in place when initiating an expanded accountability system.

Not only are these six dilemmas difficult to resolve, policymakers will need to return to them repeatedly as the frontiers of technical knowledge about indicators improves and political preferences shift. States, districts, and individual schools will have to seek consensus on basic political questions of who is accountable for which outcomes, and they will also have to find new ways to use information about schooling effectively.

Because the use of performance data from accountability systems has been neglected traditionally by States and local districts, the OERI Study Group focused most of its recommendations on that issue. These recommendations are summarized in the next section.

Section 6

The Study Group's Recommendations for Improving The Design and Use of Performance Accountability Systems

The OERI Study Group holds the firm belief that State accountability systems cannot realize their full potential as policy instruments unless they can be used to improve schools. Reporting to the public is a critical mechanism for ensuring that schools are answerable for their performance. But if the use of accountability data stops with newspaper articles or reports to parents, attempts to remedy poor performance or sustain high achievement may be haphazard, ineffective, and short-lived. Accountability data should be collected and used in ways that reflect and inform the entire schooling process—how schools are organized, what is taught, how it is taught, and what is learned.

The Study Group also recognizes that each State must devise a system that meets its own needs and political traditions. Consequently, the recommendations offered here provide general guidelines for States to consider in strengthening the link between data use and improved schooling. The guidelines accommodate both the diversity and complexity of State accountability systems.

The Study Group's recommendations fall into three categories:

1. General suggestions for designing accountability systems that serve all students;

2. Ways to incorporate accountability systems into a strategy for improving low achieving schools; and

3. Ways to sustain high achieving or significantly improving schools.

Appendix B presents a list of people whom States can call upon in addressing many of the challenges associated with designing and using a performance accountability system.

Designing Better Accountability Systems

Make the purpose of accountability systems clear. States vary considerably in what they expect their accountability systems to accomplish, but whatever those purposes are, they must be communicated clearly. One source of variation is the decision about who is held responsible. Some States have decided that local districts and schools should be held formally accountable for students' mastering a set of basic skills but have left the scope and content of curricula entirely to local discretion as long as at least those basic skills are covered. In an effort to make curricula more uniform across districts, policymakers in other States have decided that the State should establish a set of frameworks specifying

a broad range of curricula content and then reflect those specifications in its tests. Other differences in purpose stem from whether or not States decide to link policy actions to their testing programs and what those actions are. However, whatever the State's purposes may be, they must be communicated to principals and teachers. Teachers' concerns about "teaching to the test" could be greatly assuaged if teachers knew exactly what the State expects testing programs to accomplish.

The purpose of a State accountability system will be much clearer if policymakers, educators, and the public carefully consider the uses they want such a system to serve before they begin modifying or expanding the one currently in place. It is necessary to decide what is to be done with data once they are collected (for example, use them for purely informational purposes, link them to State or local policy, use them to inform school practice). It is also necessary to consider what type of system is most compatible with social and political conditions in a particular State. Finally, once consensus is reached about the system's purpose, sufficient development time should be allowed to design the accountability system and its component indicators to be consistent with its intended use.

Include multiple indicators. Whether an accountability system is used to reward high performance, remedy poor performance, or inform an overall school improvement process, it should collect data on multiple indicators to provide an integrated picture of the schooling environment. Student-related indicators might include:

- Student course-taking and, for elementary students, the amount of time spent on different subjects;
- Attendance;
- Promotion and dropout rates;
- Proportion of vocational education students meeting academic requirements;
- Proportion of students meeting university entrance requirements;
- Student participation in the arts and extracurricular activities;
- Student attitudes and social behavior (for example, vandalism); and
- Student achievement, broadly defined to include writing skills, higher order skills, and subject matter competence.

These data should be be broken down by sex, ethnicity, and socioeconomic status so that performance can be monitored for different groups of students. Despite difficulties in collecting data on students' socioeconomic status, the Study Group recommends that States and local districts seek valid proxies and collect such data.

To provide a comprehensive picture of the schooling context, States should also consider collecting data on

- Teacher quality;
- Fiscal resources and their allocation;
- Administrative leadership;
- Curricula;
- Local assessment, evaluation, and planning systems; and
- Community support.

Select indicators that are aligned with State or district goals and are not limited by today's tests and measures. To strengthen the connection between accountability data and school improvement activities, accountability systems should be linked to State and local educational goals and to current programs for achieving those goals. Citizens and parents, as well as educators, should be involved in establishing performance and improvement goals.

Policymakers also need to base the design of an accountability system on an integrated model of schooling. Such a model can provide clear guidelines for adding new indicators and replacing others as measurement technology improves (for example, with the development of better tests of higher order skills or measures of teacher quality) or as State and district goals shift. Accountability systems need to be dynamic in responding to new policy priorities and in applying new technologies, but States must make certain that in changing individual indicators, their systems still measure the central features of schooling in consistent and meaningful ways.

Include those indicators called for by the Council of Chief State School Officers if possible. The CCSSO indicators (listed in appendix C) include measures of student outcomes, school context, and educational policies and practices. The CCSSO plans to report data on these indicators for all States by 1990, as part of its continuing effort to design high-quality educational measures that are comparable across States.

Make school- and district-level accountability data public and tailor performance reports to different audiences. Public disclosure lies at the heart of democratic accountability, whether the news is good or bad. But disclosure means more than just publishing data; it also requires making performance reports understandable to the various publics concerned about schools—policymakers, educators, parents, the media, and the general public.

Effective public disclosure requires devoting substantial time and energy to placing raw data in context and helping various audiences to interpret them. Accountability systems do not produce unequivocal information about the condition of schooling, and most of the data will be susceptible to various interpretations. By thoughtfully analyzing and reporting the full range of responsible interpretations, States and local districts can help their various audiences avoid jumping to hasty or simplistic explanations and conclusions and can encourage real school improvement.

Ensure that local districts and schools find some use for accountability data, beyond State rewards or punishments. States, districts, and individual schools have very different information needs. Information sufficient to determine whether State or local performance accountability standards are being met is often considerably less detailed than the data schools and teachers need to improve programs or diagnose the instructional needs of individual students. However, to the extent that State accountability requirements are consistent and coordinated with local data needs, the system will be less burdensome and more useful to principals

and teachers. Local districts and schools need to feel that the accountability system is not an unwelcome or irrelevant intrusion from the State, but is useful for informing their own decisions and activities.

Establish an independent oversight body to monitor the development and implementation of the performance accountability system. This body should include policymakers, educators, educational testing and policy experts, and representatives of the public. Its functions should be to communicate the purposes and rationale for the system clearly, monitor its technical quality, including student outcome measures; to assess its effect on school practice and curricula; and to recommend needed changes to the legislature or State board of education. Such a body is particularly important in view of the likely changes in testing and measurement technology that will occur over the next decade and the shifts in priorities as accountability systems become a more central part of the educational policy landscape.

Using Accountability Data To Improve Low Performing Schools

Use multiple criteria to identify low performing schools and districts. Performance should be judged both in absolute terms, such as meeting a defined standard, and relatively, by gauging change over time (for example, the extent to which a school or district has improved relative to its past performance). Particular attention should be paid to the distribution of performance across different schools within a district and different types of students within a school. Once districts or schools are designated as low performing, additional data that can inform the improvement process should be gathered from on-site reviews. However, the actual designation should be based on statistical indicator data, because such data are viewed as comparable and reliable across districts and thus less subject to political pressure.

Provide sufficient opportunities for locally directed improvement activities before initiating State intervention; if State intervention is considered appropriate, it should proceed in well-defined stages. Some States find any type of State activity beyond technical assistance or financial aid inappropriate, given their political traditions. Others, however, consider State-mandated improvement activities and even intervention to be proper. No matter what role is considered appropriate, locals should have ample opportunity to improve their own performance, and the State's improvement activities should move through progressive steps:

1. Inform local districts and schools of problems that are identified, so that educators and boards of education can take action;

2. Inform citizens so that they can take action;

3. Target additional resources and assistance;

4. Mandate improvement activities (for example, require districts to develop a remediation plan and then monitor their progress on it or withhold State funds or require increased local fiscal effort); and

5. Begin State intervention (for example, direct involvement in district management, including the removal of local administrators).

Some States may find the first two or three activities desirable; others may find some of the activities in the fourth category suited to their needs and strategies; still others may find direct intervention appropriate as a last step. However, no matter how far a State moves on this continuum, it should make certain that it moves sequentially and gives local districts adequate opportunity to respond at each step.

Using Accountability Data To Sustain High Performing or Encourage Significantly Improving Schools

Each State must develop its own definition of what constitutes high performance, but in developing such a definition, it should make certain that a common understanding exists among all relevant parties (policymakers, educators, parents, and students). Policy action targeted at high performing schools is intended to maintain and encourage that performance. However, such achievement may be defined in a variety of ways: against an absolute standard, in comparison with similar schools, or in terms of significant improvement over time. States may wish to base an incentive program on any or all of these notions of high performance, but whichever they choose, the selection criteria must be well understood. This is particularly true if the State decides to reward significantly improving schools. Members of the Study Group caution that the public needs a clear understanding of selection criteria that designate a school with low absolute test scores as "high performing" when scores improve significantly. Therefore, it is critical that notions of improvement and gain be clearly stated if they are used as criteria.

Any identification of high performing or significantly improving schools should be objective and based on multiple indicators, not on a single test score. Indicators of special relevance for high performing schools might include:

- Measures of higher order skills and writing ability;
- Course enrollments, including the proportion enrolled in advanced placement classes;
- The proportion of students meeting university entrance requirements;
- The proportion of vocational students meeting academic standards; and
- The distribution of performance among varying student groups.

As with all indicators of school performance, the State should take into consideration how different types of students are achieving on each of these indicators. The Study Group also recommends that States may want to consider using additional verification from other sources such as site visits.

If State recognition of performance rests upon an application process, the State (or another agency) should offer technical assistance to districts or schools that require help in applying. Such assistance may be particularly important in States with locally defined goals and might be provided by intermediate service units, universities, or other organizations close to the local level.

Consider a broad range of programs and incentives. Among the strategies that might be used to sustain high performance are cash awards, less frequent State or local monitoring, release from some State regulations, and resources to use in assisting other schools. All of these approaches require careful consideration. For example, cash awards can be problematic if they are too high (creating serious inequities among schools) or too low (providing an insufficient incentive). The State also has to decide how award money can be spent. In some States, incentive grants can be used only to supplement the school budget; in others, funds can be allocated as salary bonuses to school personnel. Similarly, while the provision of resources to high performing schools to assist less successful ones may create useful models, States also need to ensure that this process reflects knowledge about successful strategies for change and innovation.

Afterword

Making Performance Accountability Systems Responsible and Responsive

The experience of schools and districts across the Nation clearly demonstrates that accountability systems are powerful tools. Teachers, administrators, and local communities respond to the data such systems produce, particularly to the implicit message embedded in the nature of current outcome measures. Like any other powerful tool, these systems are as good or as bad as their design. In their current, rather primitive state, they appear to produce some unintended effects that may constrain instruction and shape administrative policies in ways that many find inappropriate.

But performance accountability systems are dynamic. They will change—and must change—as indicator technology evolves and as the political context shifts. Intelligent efforts to improve accountability systems will help to ensure that they will evolve in ways that harness the systems appropriately to improvement strategies for American schools in all their diversity.

64

References

Calfee, R. (1988). *Indicators of literacy*. Santa Monica, CA: The RAND Corporation, Center for Policy Research in Education.

California State Department of Education. (1987). *Performance report for California schools.*

Cannell, J. J. (1987). *Nationally normed elementary achievement testing in America's public schools. How all fifty states are above the national average.* Daniels, WV: Friends For Education.

Congressional Budget Office. (1987, August). *Educational achievement: Explanations and implications of recent trends.* Washington, DC: U.S. Government Printing Office.

Connecticut State Board of Education. (1986). *Condition of education in Connecticut, elementary and secondary. Meeting the challenge.*

de Neufville, J. I. (1985). *Social indicators and public policy.* New York: American Elsevier.

Illinois State Board of Education. (1987). *Report card.* (sample).

MacRae, D., Jr. (1985). *Policy indicators: Links between social science and public debate.* Chapel Hill, NC: The University of North Carolina Press.

Murnane, R. J. (1987). Improving education indicators and economic indicators: The same problems? *Educational Evaluation and Policy Analysis, 9* (2), 101-116.

Planning, evaluation, reporting. (1987, October). Annual Report, District 742 Community Schools, vol. 4, St. Cloud, MN.

Savage, D. G. (1984, April 16). California basic skills test still trying to prove itself. *Los Angeles Times*, p. 1.

Savage, D. G. (1984, April 15). U.S. schools still lack a yardstick. *Los Angeles Times*, p. 1.

Shavelson, R., McDonnell, L., Oakes, J., Carey, N., & Picus, L. (1987). *Indicator systems for monitoring mathematics and science education.* Santa Monica, CA: The RAND Corporation.

South Carolina Department of Education. (1987, November). *School performance report 1986-87.* Office of School District Accreditation and Assessment.

Acknowledgments

The creation and operation of the OERI State Accountability Study Group were overseen by the OERI Task Force on Accountability and Autonomy in State Educational Reform chaired by Emily Wurtz and composed of John Coulson, James Fox, Gail Gordon, Charles Haughey, Deborah Inman, Jacqueline Jenkins, Hunter Moorman, Martin Orland, Paul Planchon, Doris Redfield, and Adria White. Leadership and support were given by Chester E. Finn, Jr., Assistant Secretary of OERI; Milton Goldberg, Director of Programs for the Improvement of Practice; David Mack, Director of the Educational Networks Division; and Gordon McAndrew, Team Leader.

Terry Peterson was the chairperson of the Study Group, helping to plan and preside over its meetings. Lorraine McDonnell and Jeannie Oakes wrote the report, organizing information given by the Study Group members with research done by the Center for Policy Research in Education and others. Study Group members (listed in appendix A) shared information about their States' practices and success and formulated the Study Group recommendations. Ramsay Selden collected information on accountability practices among the 50 States through the Council of Chief State School Officers Assessment Center's 50 State liaisons.

The report was edited by Cynthia Dorfman with invaluable suggestions from Kate Dorrell and Phil Carr.

Appendix A

STATE ACCOUNTABILITY STUDY GROUP MEMBERS

CALIFORNIA

Alex Law, 9045 Laguna Lake Way, Elk Grove, CA 95624

Rick Pratt, Legislative Analyst's Office, 925 "L" Street, Rm. 1000, Sacramento, CA 95814, (916) 445-8641

COLORADO

Judy Burnes, Supervisor of Planning and Evaluation Unit, Colorado Department of Education, 201 East Colfax, Denver, CO 80203, (303) 866-6824

Ann Kraetzer, Department of Planning and Evaluation, Colorado Springs Public Schools, 1115 North El Paso St., Colorado Springs, CO 80903, (303) 520-2000

CONNECTICUT

Pat Forgione and **Peter M. Prowda**, State Department of Education, P.O. Box 2219, Hartford, CT 06145, (203) 566-3577; 566-7117

D'Ann Mazzocca, Senior Research Analyst, Office of Legislative Research, 18-20 Trinity, Hartford, CT 06106, (203) 566-8400

FLORIDA

Thomas H. Fisher, Administrator, Department of Education, Assessment, Testing and Evaluation Section, Knott Bldg., Rm. 506, Tallahassee, FL 32399, (904) 488-8198

Ray Turner, Assistant Superintendent of Accountability, Dade County Public Schools, 1450 NE 2nd Ave., Miami, FL 33132, (305) 376-1506

GEORGIA

Joy Berry, Executive Director, Human Relations Commission of Georgia, 270 Washington St., SW, Room 603, Atlanta, GA 30334, (404) 651-9115

Larry Gess, Director of Strategic Planning, Georgia State Department of Education, 1758 Twin Towers, Atlanta, GA 30334, (404) 656-0352

ILLINOIS

Mary Lou Cowlishaw, State Representative, 552 South Washington St., Suite 119, Naperville, IL 60540, (217) 782-6507

Gail Lieberman, Assistant to the Governor for Education, Governor's Office, 2-1/2 State House, Springfield, IL 62706, (217) 782-4921

Sally Bulkley Pancrazio, Chief, Research and Evaluation, Illinois State Board of Education, 100 North First Street, Springfield, IL 62777, (217) 782-3950

MASSACHUSETTS

Alan Hartman, Massachusetts Department of Education, 1385 Hancock Street, Quincy, MA 02169, (617) 770-7204

Richard Kraus, State Senator, Co-Chair Education Committee, State House, Room 511, Boston, MA 02133, (617) 722-1432

Robert Schwartz, Special Assistant to the Governor for Education, State House, Room 173, Boston, MA 02133, (617) 727-0770

MINNESOTA

David Bennett, Superintendent of Schools, 360 Colborne, St. Paul, MN 55102, (612) 293-5150

Ken Kelsey, Chairperson, State Curriculum Advisory Committee, 5835 Sandy Drive, St. Cloud, MN 56301, (612) 255-3007

James Pehler, State Senate, Room 9, State Capitol, St. Paul, MN 55155, (612) 296-4241

NEW JERSEY

Valarie French, Special Assistant to the Commissioner of Education, Department of Education, 225 W. State Street, CN 500, Trenton, NJ 08625, (609) 292-4450

Richard Mills, Commissioner of Education, Department of Education, Montpelier, VT 05602, (802) 828-3135

SOUTH CAROLINA

Terry Peterson, Winthrop College and Joint Business-Education Legislative Committee, 1122 Lady St., Rm. 1005, Columbia, SC 29201, (803) 734-0489

Mike Turner, Director, Office of School District, Accreditation and Assessment, SC State Department of Education, 701 Rutledge Building, Columbia, SC 29201, (803) 734-8536

NATIONAL GOVERNORS' ASSOCIATION

Michael Cohen, National Governors' Association, 444 North Capitol Street, Washington, DC 20202, (202) 624-7815

Susan Traiman, National Governors' Association, 444 North Capitol Street, Washington, DC 20202, (202) 624-5385

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

Ramsay Selden, Council of Chief State School Officers, 400 North Capitol Street, Suite 377, Washington, DC 20001, (202) 624-7700

NATIONAL ASSOCIATION OF STATE BOARDS OF EDUCATION

Janice Earle, National Association of State Boards of Education, 701 North Fairfax Street, Suite 340, Alexandria, VA 22314, (703) 684-4000

NATIONAL CONFERENCE OF STATE LEGISLATURES

Mark Weston, National Conference of State Legislatures, 150 17th Street, Suite 2100, Denver, CO 80263, (303) 623-7800

RAND CORPORATION

Lorraine McDonnell and **Jeannie Oakes,** RAND Corporation, 1700 Main Street, Santa Monica, CA 90406, (213) 393-0411

OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT

Emily O. Wurtz, OERI, 555 New Jersey Ave., NW, Washington, DC 20202, (202) 357-6172

Appendix B

INFORMATION SOURCES

For further information about the design and use of educational indicators, we recommend the following publications:

Jane L. David, 1987. *Improving Education with Locally Developed Indicators.* New Brunswick, NJ: Center for Policy Research in Education/Rutgers University.

Richard J. Murnane and **Senta A. Raizen,** eds. 1988. *Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12.* Washington, DC: National Academy Press.

Jeannie Oakes, 1986. *Educational Indicators: A Guide for Policymakers.* Santa Monica, CA: Center for Policy Research in Education/The RAND Corporation.

Representatives from each of the 10 States participating in the OERI Study Group prepared a detailed description of their States' accountability systems. These reports can be obtained from:

Emily O. Wurtz, Office of Educational Research and Improvement, U.S. Department of Education, 555 New Jersey Ave., NW, Washington, DC 20202

The people and organizations listed below know and have experience with different aspects of developing indicators and accountability systems. They may be contacted for further information.

DEVELOPING MULTIPLE INDICATORS

Eva Baker, Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, California 90024

Jeannie Oakes, The RAND Corporation, 1700 Main St., Santa Monica, California 90406

Martin Orland, Office of Educational Research and Improvement, U. S. Department of Education, 555 New Jersey Ave., NW, Washington, DC 20202

Senta A. Raizen, Center for Improving Science Education, 290 South Main St., Andover, Massachusetts 01810

Ramsay Selden, Council of Chief State School Officers, Assessment Office, 400 North Capitol Street, Suite 377, Washington, DC 20001

ALTERNATIVES TO TRADITIONAL STANDARDIZED TESTS

Joan Baron, Connecticut State Department of Education, P.O. Box 2219, Hartford, CT 06145

Rexford Brown, Education Commission of the States, 1860 Lincoln Street, Suite 300, Denver, CO 80295,

Leigh Burstein, Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, California 90024

Robert Calfee, Graduate School of Education, Stanford University, Stanford, California 94305

Norman Fredrickson, Director of Psychological Studies, Educational Testing Service, P.O. Box 2923, Princeton, NJ 08541

William McMillan, Director of Assessment, Minnesota State Department of Education, Capitol Square Building, 550 Cedar St, St. Paul, Minnesota 55101

Ina Mullis, National Assessment of Educational Progress, Educational Testing Service, P.O. Box 2923, Princeton, NJ 08541

IMPLEMENTING ACCOUNTABILITY SYSTEMS

Michael Cohen, National Governors' Association, Hall of the States, 444 North Capitol Street, Washington, DC 20001

Pat Forgione, Connecticut State Department of Education, P.O. Box 2219, Hartford, CT 06145

Larry Gess, Director of Strategic Planning, Georgia State Department of Education, 1758 Twin Towers, Atlanta, GA 30334

Alex Law, 9045 Laguna Lake Way, Elk Grove, CA 95624

Lorraine McDonnell, The RAND Corporation, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406

Ramsay Selden, Council of Chief State School Officers, Assessment Office, 400 North Capitol Street, Suite 377, Washington, DC 20001

USING ASSESSMENT INFORMATION FOR LOCAL SCHOOL IMPROVEMENT

Leigh Burstein, Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles, California 90024

Robert Calfee, Graduate School of Education, Stanford University, Stanford, California 94305

Jane L. David, 3144 David Avenue, Palo Alto, CA 94303

Ken Kelsey, Chairperson, State Curriculum Advisory Committee, 5835 Sandy Drive, St. Cloud, MN 56301

Ann Kraetzer, Department of Planning and Evaluation, Colorado Springs Public Schools, 1115 North El Paso St., Colorado Springs, CO 80903

Andrew Porter, Institute for Research on Teaching, Michigan State University, Erickson Hall, East Lansing, MI 48824

REPORTING ACCOUNTABILTY DATA

Alex Law, 9045 Laguna Lake Way, Elk Grove, CA 95624

John May, South Carolina Department of Education, 701 Rutledge Building, Columbia, SC 29201

Sally Bulkley Pancrazio, Research and Evaluation, Illinois State Department of Education, 100 N. Front Street, Springfield, IL 62777

Peter M. Prowda, Connecticut State Department of Education, P.O. Box 2219, Hartford, CT 06145

Ramsay Selden, Council of Chief State School Officers, Assessment Office, 400 North Capitol Street, Suite 377, Washington, DC 20001

COMPARING SCHOOL PERFORMANCE

William Cody, Council of Chief State School Officers, Assessment Office, 400 North Capitol Street, Suite 377, Washington, DC 20001

Mark Fetler, Division of Program Evaluation and Research, California State Department of Education, 721 Capitol Mall, 4th Floor, Sacramento, CA 95814

Michael Turner, Accreditation and Assessment, South Carolina State Department of Education, 701 Rutledge Building, Columbia, SC 29201

Ray Turner, Assistant Superintendent of Educational Accountability, Dade County Public Schools, 1450 NE 2nd Ave., Miami, FL 33132

INCENTIVES, REWARDS, AND SANCTIONS

David Bennett, St. Paul Schools, 360 Colborne, St. Paul, MN 55102

Saul Cooperman and **Valarie French**, New Jersey State Department of Education, 225 W. State Street, CN 500, Trenton, NJ 08625

John May, South Carolina Department of Education, 701 Rutledge Building, Columbia, SC 29201.

NATIONAL PERSPECTIVES ON STATE ACCOUNTABILITY POLICIES

Michael Cohen and **Susan Traiman,** National Governors' Association, Hall of the States, 444 North Capitol Street, Washington, DC 20001

Terry Peterson, Winthrop College, 1122 Lady St., Room 1005, Columbia, SC 29101, (803) 734-0489

Chris Pipho, Education Commission of the States, 1860 Lincoln Street, Denver, CO 80295

Ramsay Selden, Council of Chief State School Officers, Assessment Office, 400 North Capitol Street, Suite 377, Washington, DC 20001

Mark Weston, National Conference of State Legislatures, 150 17th Street, Suite 2100, Denver, CO 80265

Appendix C

STATUS OF COUNCIL OF CHIEF STATE SCHOOL OFFICERS EDUCATIONAL INDICATORS

Table C.1--EDUCATIONAL INDICATORS ADDRESSING *OUTCOMES*

Indicators	Measures	Status	Target Date
Attendance Rates	Percent Attend/Membership	Under development	1988
Completion Rates	HS Graduates	Recommended	1987
	GEDs Awarded	Under development	1987
Achievement Levels	Performance in Reading, Math, Science, Social Studies, and English at the Elementary, Intermediate and Secondary Levels	Under development	1990
Post-School Outcomes	Job Placement	To be developed	1989
	Postsecondary Schooling	Available	1987
	Military	Available	1987
	Civic Involvement	To be developed	1989
	Attitudes		
	Student Followup	To be developed	1989
	Parents and Employers	To be developed	1989

Table C.2—EDUCATIONAL INDICATORS ADDRESSING *CONTEXT*

Indicators	Measures	Status	Target Date
School System	Total School-Age Population	Available	1987
	Number and Size Ranges of School Districts	Available	1987
	Percent School-Age Population Urban/Suburban/Rural	Available	1987
Population	Educational Attainment	Available	1987
	Per Capita Income	Available	1987
	Percent Voting	Available	1987
Resources	Per Pupil Wealth	Available	1987
	Pre-School Participation	To be developed	1990
Student Needs	Percent Disadvantaged	Available	1987
	Percent Handicapped	Recommended	1988
	Percent Limited English	Available	1987
	Percent Private and Home Schooling	To be developed	1990

Table C.3—INDICATORS ADDRESSING EDUCATIONAL *POLICIES AND PRACTICES*

Indicators	Measures	Status	Target Date
INSTRUCTION			
Instructional Time	Days/Years	Under development	1987
	Hours/Day	Under development	1987
	Percent Engaged Time	To be developed	1990
	Time Allocated To: Reading & Math, Science, Social Studies & Arts	Under development	1990
Instructional Program	Kindergarten Required	Under development	1987
	Graduation Requirements	Under development	1987
	Honors and AP Course Offerings	Under development	1988
Effective Schooling	Instructional Leadership	To be developed	1987
	Instructional Environment	To be developed	1987
	Home and Community Involvement	To be developed	1987
TEACHER QUALITY			
Teacher Preparation	Course Units	Under development	1987
	Testing	Under development	1987
	Practice Teaching	Under development	1987
	Inservice Programs	Under development	1987
Teacher Work Force	Percent Advanced Degrees	Under development	1988
	Experience in Years	Under development	1988
	Classroom Performance	To be developed	1990
RESOURCE ALLOCATION			
Expenditures	Total Per Pupil	Under development	1988
	Teacher Salaries	Under development	1988
	Instructional Materials	Under development	1988
Effort	Expenditures for Educ./Other Gov't. Expenditures	Available	1987
PARTICIPATION POLICIES			
	Compulsory Age	Under development	1987